CRM 90-201 / February 1991

A Comparison of Hands-On and Job-Knowledge Tests: implications for Better Test Development

Neil B. Carey

CNA
CENTER FOR NAVAL ANALYSES
4401 Ford Avenue • Post Office Box 16268 • Alexandria, Virginia 22302-0268

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

Work conducted under contract N00014-91-C-0002.

REPORT DOCUMENTATION PAGE

Form Approved
OPM No. 0704-0188

1. AGENCY USE ONLY *(Leave Blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	February 1991	Final

4. TITLE AND SUBTITLE

A Comparison of Hands-On and Job-Knowledge Tests: Implications for Better Test Development

5. FUNDING NUMBERS

C - N00014-91-C-0002

PE - 65153M

PR - C0031

6. AUTHOR(S)
Neil B. Carey

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Center for Naval Analyses
4401 Ford Avenue
Alexandria, Virginia 22302-0268

8. PERFORMING ORGANIZATION REPORT NUMBER

CRM 90-201

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Commanding General
Marine Corps Combat Development Command (WF 13F)
Studies and Analyses Branch
Quantico, Virginia 22134

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Hands-on performance tests are the benchmark against which other measures of proficiency shoudl be compared. However, hands-on performance tests are expensive, time-consuming, and sometimes dangerous to personnel or equipment. This paper analyzes the relationship between hands-on performance tests and job-knowledge tests to propose better methods for developing job-knowledge tests and to determine when job-knowledge tests could best be used in place of hands-on performance tests.

14. SUBJECT TERMS

Aptitude tests, Comparison, Infantry personnel, Job training, Marine Corps personnel, Performance (human), Performance tests, Proficiency, Reliability, Statistical processes, Statistical samples, Test methods

15. NUMBER OF PAGES
89

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
CPR	CPR	CPR	SAR

CNA CENTER FOR NAVAL ANALYSES
4401 Ford Avenue • Post Office Box 16268 • Alexandria, Virginia 22302-0268 • (703) 824-2000

11 March 1991

MEMORANDUM FOR DISTRIBUTION LIST

Subj: Center for Naval Analyses Research Memorandum 90-201

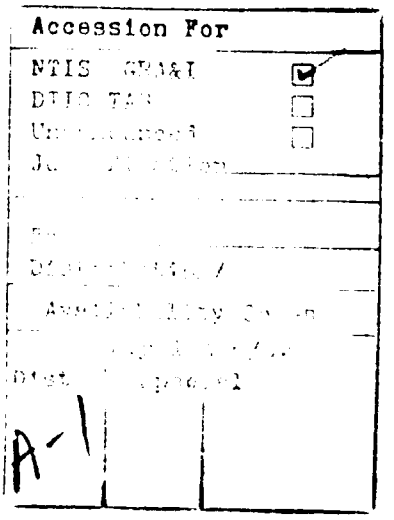
Encl: (1) CNA Research Memorandum 90-201, *A Comparison of Hands-On and Job-Knowledge Tests: Implications for Better Test Development*, by Neil B. Carey, Feb 1991

1. Enclosure (1) is forwarded as a matter of possible interest.

2. Hands-on performance tests are the benchmark against which other measures of proficiency should be compared. However, hands-on performance tests are expensive, time-consuming, and sometimes dangerous to personnel or equipment. This paper analyzes the relationship between hands-on performance tests and job-knowledge tests to propose better methods for developing job-knowledge tests and to determine when job-knowledge tests could best be used in place of hands-on performance tests.

Lewis R. Cabe
Director
Manpower and Training Program

Distribution List:
Reverse page



Subj: Center for Naval Analyses Research Memorandum 90-201

Distribution List

SNDL

45A2	CG I MEF
45A2	CG II MEF
45A2	CG III MEF
45B	CG FIRST MARDIV
45B	CG SECOND MARDIV
A1	ASSTSECNAV MRA
A1	DASN MANPOWER (2 copies)
A2A	CNR
A6	HQMC MPR & RA
	Attn: Code M
	Attn: Code MR
	Attn: Code MP
	Attn: Code MM
	Attn: Code MA (3 copies)
A6	CG MCRDAC, Washington
A6	HQMC AVN
FF38	USNA
	Attn: Nimitz Library
FF42	NAVPGSCOL
FF44	NAVWARCOL
FJA1	COMNAVMILPERSCOM
FJB1	COMNAVCRUITCOM
FJA13	NAVPERSRANDCEN
	Attn: Technical Director (Code 01)
	Attn: Technical Library
	Attn: Director, Manpower Systems (Code 61)
	Attn: Director, Personnel Systems (Code 62)
FT1	CNET
V8	CG MCRD Parris Island
V8	CG MCRD San Diego
V12	MCCDC
	Attn: Studies and Analyses Branch
	Attn: Director, Warfighting Center
	Attn: Warfighting Center, MAGTF Proponency and Requirements Branch (2 copies)
	Attn: Director, Training and Education Center
V12	CG MCRDAC, Quantico

OPNAV

OP-01
OP-11
OP-12
OP-13

OTHER

Defense Advisory Committee on Military Personnel Testing (8 copies)
Joint Service Job Performance Measurement Working Group (13 copies)

CRM 90-201 / February 1991

A Comparison of Hands-On and Job-Knowledge Tests: Implications for Better Test Development

Neil B. Carey

Operations and Support Division

CNA
CENTER FOR NAVAL ANALYSES
4401 Ford Avenue • Post Office Box 16268 • Alexandria, Virginia 22302-0268

ABSTRACT

Hands-on performance tests are the benchmark against which other measures of proficiency should be compared. However, hands-on performance tests are expensive, time-consuming, and sometimes dangerous to personnel or equipment. This paper analyzes the relationship between hands-on performance tests and job-knowledge tests to propose better methods for developing job-knowledge tests and to determine when job-knowledge tests could best be used in place of hands-on performance tests.

EXECUTIVE SUMMARY

The Marine Corps must periodically assess the ability of its personnel to perform mission-related tasks. The best criterion of mission-related proficiency would be a hands-on performance test (HOPT), because of its validity and objectivity. HOPTs have been established as the benchmarks against which other measures should be compared. However, HOPTs are expensive, time-consuming, and sometimes dangerous to personnel or equipment.

In certain circumstances, job-knowledge tests (JKTs) can be inexpensive substitutes (i.e., surrogates) for hands-on performance tests. Previous research shows that JKTs show promise as surrogates for diagnosis of some training needs, but these studies do not fully analyze the strengths and weaknesses of job-knowledge tests.

The current study extends previous studies to indicate the conditions under which job-knowledge tests should be used as surrogates and to provide guidance for developing better job-knowledge tests.

Both HOPTs and JKTs measure job proficiency with some degree of error. Ideally, scores on the HOPTs and JKTs should be the same; that is, JKT items should have the same difficulty level as HOPT tasks. Such equality should be present for the total test score and extend to scores at the duty-area level. Duty-area scores provide informative comparisons that may be useful for examining training needs.

Discrepancies between HOPTs and JKTs should be minimized, because discrepancies can lead to misinterpretation of training needs. This research memorandum uses "sample-difficulty" analyses to quantify causes of HOPT-JKT discrepancies. These analyses postulate HOPT-JKT discrepancies to be the result of the following two factors, plus error:

$$\Delta_{HOPT - JKT} = \Delta_{SAMPLING} + \Delta_{DIFFICULTY} + \Delta_{ERROR}$$

On the right side of the equation, the first term indicates discrepancies that result from item sampling. These occur when the proportion of items differs from the proportion of tasks within a duty area. For example, sampling discrepancies arise when there are no JKT items to represent a HOPT task, or when there is an overabundance of items representing a single task.

The second term refers to differences in the difficulties of JKT items compared to HOPT tasks. "Difficulty" is simply the average proportion of HOPT steps performed correctly or the average fraction of examinees who respond correctly to JKT items. The difficulty component compares a task with the items written to represent that task. The "error" term represents all other differences that cannot be explained by the two other components. Sampling discrepancies could be corrected

by allocating items to be proportional to the number of HOPT tasks; difficulty discrepancies could, in many cases, be corrected by better item writing. In some cases, difficulty discrepancies are an unavoidable result of differences in the skills required by hands-on as opposed to paper-and-pencil tests.

This study analyzed data from more than 1,900 first-term Marine infantrymen. Sample-difficulty (S-D) analyses found that the most common and significant reasons for HOPT-JKT discrepancies were differences in difficulty. Further analyses were therefore focused on the quality of item writing.

Analyses of JKT item-total correlations determined that about 19 of the 150 core job-knowledge items lacked proper measurement properties. Deletion of these items improved overall HOPT-JKT correspondence approximately 15-20 percent.

It often occurred that job-knowledge items were more difficult than the actual hands-on task. This occurred when items asked for knowledge that was not required to perform the task, or items were geared for leadership functions not yet encountered by first-term Marines. Simpler item formats and better task analyses are recommended to avoid these problems with job-knowledge tests.

Conversely, it was found that traditional multiple-choice job-knowledge items could be easier than the corresponding tasks if the items reduce complex activities to a simple choice between alternatives. For example, abilities to perform complicated activities such as building field-expedient antennas are probably better measured using formats other than traditional multiple choice.

In summary, the following steps should be taken to improve job-knowledge tests:

- The proportion of items should more nearly reflect the percentage of tasks in the duty area.
- A larger number of job-knowledge items should be written.
- Items with low item-total correlations should be revised or deleted after thorough pilot testing.
- Task analyses should determine the appropriate difficulty level of questions and determine whether alternative formats are necessary to assess each HOPT skill.
- Task analyses should determine what steps or aspects of a duty area should be assessed by means of a job-knowledge test. Critical, knowledge-dependent steps should be emphasized on a job-knowledge test.

Analyses in this paper used the HOPT as the benchmark against which the JKT should be compared--hence, error associated with the HOPT is ignored. Job-knowledge tests might be a better benchmark if the objective of testing is to determine Marines' knowledge as opposed to "can do" abilities. Nevertheless, taken as a whole, this and previous research indicate that JKTs are most appropriate for measuring proficiency in the following tasks:

- *Knowledge-driven tasks* that require memory for specific facts and attention to detail among complex alternatives. Some skills in land navigation are knowledge-driven.

- *Reading-dependent tasks* in which the JKT is a close approximation to actual job requirements.

- *Time-independent tasks* in which the actual job performance allows sufficient time to recall information. Some maintenance tasks are quite time-independent.

In contrast, physical-coordination tasks, very difficult tasks, and time-critical tasks are usually inappropriate to measure with a multiple-choice JKT. Extremely difficult tasks should not be measured by means of multiple choice, because guessing provides a lower bound on measured proficiency levels. The paper indicates that physical-coordination and time-critical tasks should be measured by HOPTs or high-fidelity simulations. Some aspects of complex construction tasks can be measured using alternative item formats. However, the critical concern is that the measurement be an accurate reflection of an individual's proficiency level and not be contaminated by reading or writing abilities that can highly influence performance on job-knowledge tests.

CONTENTS

CONTENTS (Continued)

ILLUSTRATIONS

TABLES

INTRODUCTION

Periodically, the Marine Corps must assess the ability of each of its personnel to perform mission-related tasks. This research memorandum analyzes how Marines' proficiencies could be measured more accurately. It compares the results of hands-on performance tests (HOPTs) and job-knowledge tests (JKTs), then recommends ways in which better job-knowledge tests could be developed. This research also studies the circumstances in which a JKT is most appropriate.

The best criterion for measuring relative training success would be a HOPT because of its validity and objectivity. The National Academy of Sciences Committee, which provided scientific oversight of the Job-Performance Measurement Project, has recommended use of HOPTs as the benchmarks against which other measures should be compared [1]. However, HOPTs are expensive, time-consuming, and sometimes dangerous to personnel or equipment.

In certain circumstances, JKTs can be inexpensive substitutes (i.e., surrogates) for HOPTs within a duty area. A duty area is a domain of job performance defined by Individual Training Standards, such as land navigation or tactical measures. To be used as a surrogate, a JKT should provide the same profile of duty-area strengths as would a HOPT [2,3]. In other words, the difficulty of each duty area should be the same for JKTs and HOPTs. As indicated in the quote below, the National Academy of Sciences Committee has concluded that hands-on measures of enlisted performance are the benchmark against which surrogates such as JKTs should be compared.

> ... the project consensus is that the most direct measures of job behaviors have the greatest likelihood of meeting validity requirements. Although it is not universally applicable, the hands-on job sample test is the measure of job proficiency with the greatest fidelity to actual job performance. For each military occupational specialty under study, therefore, a hands-on test will be developed. But because such tests are very expensive to administer and, for reasons of time, cost, and safety, can only sample a small number of tasks in a given MOS, an important objective of the project is to develop additional "surrogate" measures for each MOS that are cheaper and more feasible for large-scale administration. The hands-on measure will serve as the benchmark to which the surrogate measures must compare favorably if they are to be endorsed by the Joint-Service Project [1, p. 6].

Ideally, there should be correspondence between the proportion correct of the HOPT and JKT within a duty area. If the proportion correct deviates between HOPT and JKT, a discrepancy occurs. This

research analyzes the reasons for such discrepancies and proposes methods to minimize them. Previous research has shown that, in the infantry occupational field, job-knowledge tests can be useful for diagnosing training needs in some, but not all, duty areas [2, 4, 5].

The current research uses sample-difficulty (SD) analyses, modified from Cooke [6], to quantify the degree to which discrepancies are the result of two factors, plus error as follows:

$$\Delta_{HOPT - JKT} = \Delta_{SAMPLING} + \Delta_{DIFFICULTY} + \Delta_{ERROR}$$

On the right side of the equation, the first term indicates discrepancies that result from item sampling. These occur when the proportion of items differs from the proportion of tasks within a duty area. For example, sampling discrepancies arise when there are no JKT items to represent a HOPT task, or when there is an overabundance of items representing a single task.

The second term refers to differences in the difficulties of JKT items compared to HOPT tasks. "Difficulties" are simply the average proportion of HOPT steps performed correctly or the fraction of examinees who respond correctly to a JKT item. The difficulty component compares a task with the items written to represent that task. The "error" term represents all other differences that cannot be explained by the two other components.

This research next makes recommendations about better test development techniques that could minimize the observed sample and difficulty discrepancies. Additional analyses are performed to determine specific content areas that are most appropriately assessed by means of JKTs.

METHOD

Subjects

HOPTs and JKTs were administered to more than 1,900 first-term Marines in four infantry specialties. Over 1,000 riflemen, 300 machine gunners, 300 mortarmen, and 300 assaultmen took part in this research. Individuals to be tested were randomly selected from the available Marine Corps for each Military Occupational Specialty (MOS). The sample was stratified by pay grade, length of service, and educational level. Two-hundred of the riflemen were retested with the alternate form of the performance and job-knowledge tests to determine the test-retest reliability of the testing procedures. Further description of the sample is provided in reference [7].

Measures and Reliability

Hands-On Performance Tests

The first task in developing job-performance measures was to define the requirements of Marine Corps enlisted infantrymen for each MOS. Individual Training Standards (ITS) prepared by the Marine Corps were the primary source of detailed information about the tasks required in each MOS. Analyses of the ITS were conducted to ensure that tasks selected for testing would maximize the coverage of job behaviors. In this manner, hands-on test scores would generalize to the full range of infantry job requirements. Appendix A and reference [8] provide further details of the test construction procedures.

HOPTs were developed for the selected test content. These tests were reviewed by Marine Corps job experts. They were then trial-tested and improved before a large-scale tryout was conducted with more than 200 Marines. Table 1 provides an overview of the duty areas covered in the test of infantry skills, and further details regarding test development can be found in appendix A. Test-administrator training was conducted for two weeks, during which time test administrators learned to perform all tasks and to score performance according to objective criteria.

Table 1. Examples of duty areas and tests included in hands-on performance tests of infantry skills

Duty Area	Examples of Tests
Tactical measures	Call for/adjust indirect fire
Security and intelligence	Process prisoners
M16A2 service rifle	Live fire at pop-up targets
M203 grenade launcher	Prepare for firing
Hand grenades	Throw dummy grenades
Mines	Install Claymore mines
Communication	Assemble and operate radio
Land navigation	Determine location
First aid	Treat sucking chest wound
Nuclear, biological, chemical	Prepare NBC-1 report
Light antitank weapon (LAW)	Prepare to fire
Night vision	Operations inspection
Squad automatic weapon (SAW)	Fieldstrip SAW

Internal consistency reliabilities for the HOPT were fairly high overall (varying between .88 and .83, depending on the MOS) [2,7,9], but somewhat low by duty area, as shown in table 2. These reliabilities

were high enough to use the HOPT as a benchmark against which the JKT could be judged, although lack of reliability can cause HOPT-JKT discrepancies.

Table 2. Reliability estimates of the job knowledge test and HOPT by duty area

Duty area	Core JKT		HOPT	
	Form A	Form B	Form A	Form B
Land navigation	.65	.65	.77	.69
Security and intelligence	.65	.63	.36	.27
Communications	.61	.56	.70	.65
Grenade launcher	.50	.36	.31	.31
Tactical measures	.50	.68	.45	.35
LAW	.45	.46	.76	.75
SAW	.46	.61	.61	.59
NBC	.42	.37	.54	.57
Mines	.43	.35	.85	.89
Hand grenade	.15	.13	NA	NA
Night vision	.14	NA	.51	.70
First aid	.03	.20	.65	.58

NOTE: Cronbach internal consistency reliabilities for areas marked "NA" could not be computed because there was only one item on that scale.

Job-Knowledge Tests

Development procedures for the JKTs followed those for the HOPTs and are detailed elsewhere [2, appendix B]. Overall internal consistency reliabilities of the JKT were fairly high, varying from .90 to .87, depending on the MOS. Reliabilities by duty area were moderate to low (table 2). The low reliabilities, which were partly the result of having too few items for each duty area, limited HOPT-JKT correlations by duty area and prevented useful determination of each individual's duty-area strengths and weaknesses on the basis of the JKT [2]. Correlations by duty area ranged from a high of .46 for land navigation to .04 for the night-vision device (table 3). However, the low reliabilities do not necessarily limit the JKT for determining overall proficiencies [2].

Table 3. Job-knowledge test--hands-on correlations by duty area

Land navigation	.46
NBC	.28
Tactical measures	.29
Communications	.25
LAW	.23
Grenade launcher	.22
First aid	.20
Security and intelligence	.18
Mines	.16
SAW	.11
Hand grenades	.05
Night vision	.04

RESULTS

Duty-Area Analyses

Analyses of Form A and Form B HOPT-JKT correspondence are presented in figures 1 and 2. These figures show discrepancies between the average percentage of HOPT steps performed correctly and the average percentage of examinees answering JKT items correctly, by duty area. Positive values show that the JKT was more difficult[1] than the HOPT. Ideally, HOPT-JKT discrepancies would be zero--corresponding to the solid line shown in these figures. It would also be ideal for differences between forms--contrasts between figures 1 and 2--to be negligible.

Both figures show that the job-knowledge test was more difficult than the hands-on performance test for most duty areas. For Form A, the job-knowledge test was about as difficult as the HOPT for hand grenades, land navigation, and SAW; conversely, night vision and first-aid items were much more difficult than the corresponding HOPT tasks. Form B showed mines as an area where the JKT was easier than the HOPT; the JKT was much harder for the night-vision duty area.

1. "Difficulty" for the HOPT is the average percentage of task steps performed correctly; for the JKT, "difficulty" is the average percentage of examinees responding correctly to the JKT items. When the HOPT is easier than the JKT, the average percentage of HOPT-task steps performed correctly is higher than the average percentage of examinees responding correctly to the JKT items, so discrepancy values will be positive.

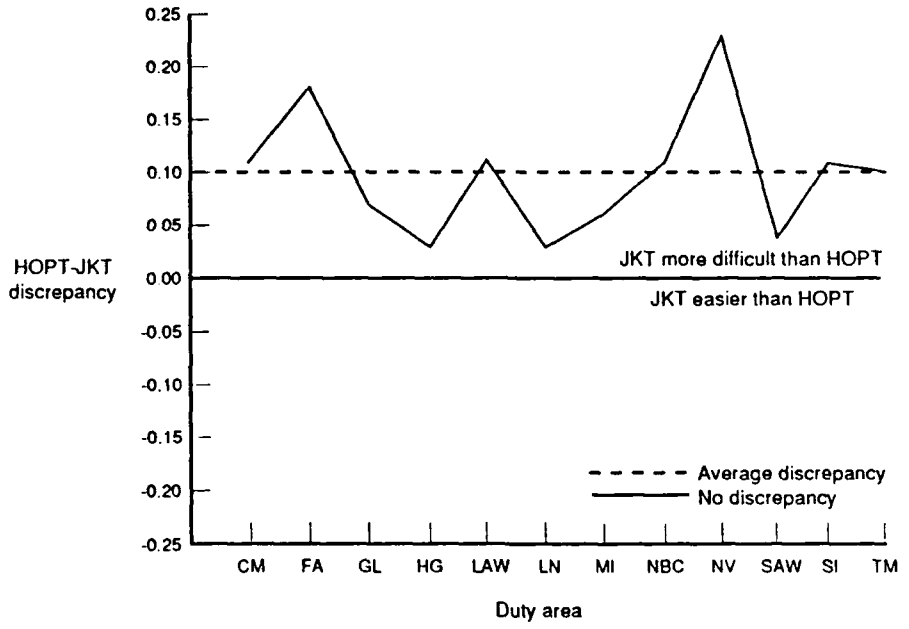


Figure 1. Overall HOPT-JKT discrepancies (Form A)

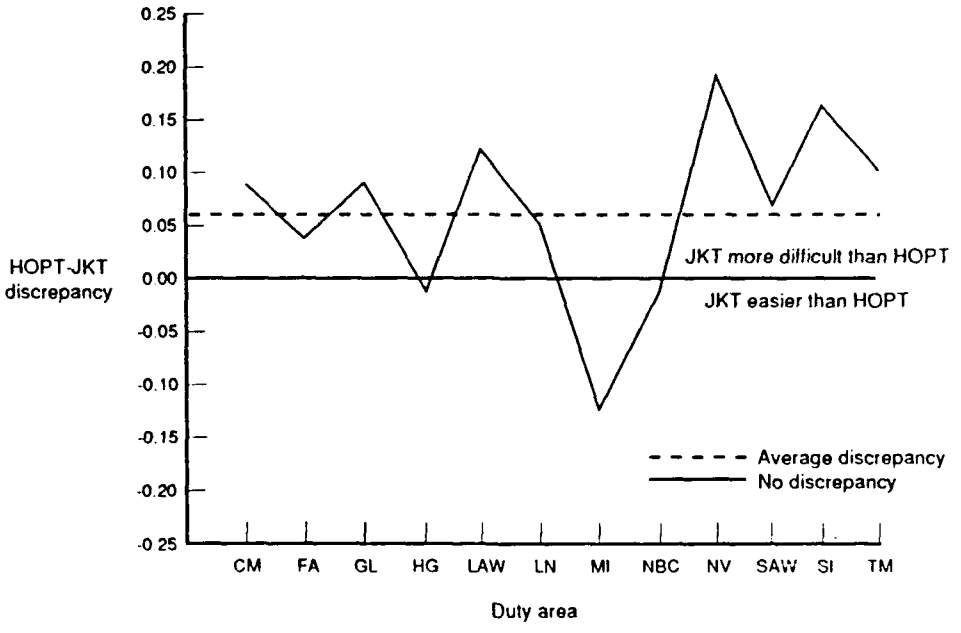


Figure 2. Overall HOPT-JKT discrepancies (Form B)

Some of these HOPT-JKT discrepancies could be attributed to unreliability, as shown in table 2. It is not surprising that the night-vision duty area, which had virtually no reliability or correlation with the HOPT, was an area where the JKT and the HOPT diverged considerably. However, communications, security and intelligence, and tactical measures also had large discrepancies despite moderate correlations with HOPTs.

For the three duty areas with large form discrepancies[1] (first aid, mines, NBC), such differences were partially the result of differences in test content. For example, replacing and recovering the Claymore mine was based on an electronic detonation for Form A versus tripwire detonation for Form B. Marines are much less proficient in emplacing and recovering Claymore mines with tripwires than with electronic devices. All other duty areas had similar test content, and no other significant form discrepancies were noted.

Duty-Area Sample-Difficulty Analyses

Two reasonable hypotheses for explaining duty-area HOPT-JKT discrepancies are that the job-knowledge test items did not provide proportional coverage of duty-area tasks ("sampling"); or the difficulty level of the JKT items was different from corresponding HOPT tasks ("difficulty").

The first source of discrepancies, "sampling," is responsible for a discrepancy when the proportions of items within a duty area differ from the proportions of tasks within a duty area. For example, sampling discrepancies arise when there are no JKT items to represent a HOPT task, or when there is an overabundance of items representing a single task. The second source, "difficulty," is responsible for a discrepancy when a JKT item is either harder or easier than the task it is supposed to represent.

The quantification of the relative importance of these two sources of discrepancies was based on "shift-share" analyses used by Cooke [6]. Appendix C explains the mathematical background of these analyses.

To begin S-D analyses, each hands-on performance task was reviewed and matched to items on the JKT to produce a "crosswalk" between the HOPT criterion and the JKT surrogate. Matches were made by reviewing each item and deciding which task it represented. Next, the reverse procedure was employed--tasks were matched to items--to confirm task-item pairings. Appendix D shows the names and average percentage

---

1. Form discrepancies are differences in the HOPT-JKT discrepancy between Form A and Form B. For example, the HOPT-JKT discrepancy for mines was about +.06 for Form A, but -.12 for Form B. These form discrepancies primarily reflect differences in test content, but they also reflect differences in the examinees who used the forms.

correct of tasks and their matching items. Note that some tasks had no corresponding items (e.g., convert azimuth), and vice-versa.

Table 4 shows the results of S-D analyses by duty area and form. The "sample" component represents the extent to which differences in the proportions of *items* within a duty area differ from the proportions of *tasks* within a duty area. It will be positive if the easier HOPT tasks constitute a larger proportion of the HOPT total score than the corresponding JKT items. For example, in Night Vision on Form A (appendix C), "visual inspection" was the easiest task (67.9 percent correct), and the second-easiest was "operations inspection" (65.3 percent correct). Each of these easier tasks accounted for 33.3 percent of the total HOPT score, but the corresponding items contributed only 25.0 percent to the total JKT score. The positive sign indicates that this sampling tended to make the total HOPT score higher than the JKT.

The "difficulty" component represents the difference in difficulty between the item and the task the item is supposed to represent. Difficulty is the average percentage of HOPT steps done correctly for a task, or the average percentage of examinees responding correctly to the matching items. The difficulty component will be positive if HOPT tasks tend to be easier than the average of corresponding JKT items. For Night Vision on Form A, all HOPT tasks were easier than the matched JKT items, so the difficulty component was positive.

In table 4, the "error" component indicates whether items discrepant in difficulty were also different in proportion sampled. In practice, this component is indistinguishable from error. It will be large and negative if the biggest difficulty differences were for tasks that also had substantial sampling discrepancies, and the sign of the discrepancies match. For Communications on Form A, the error component is large because the two easiest tasks (with large positive difficulty discrepancies) also had no matching items (i.e., a substantial positive sampling difference).

In general, table 4 indicates that difficulty differences are responsible for the single largest amount of HOPT-JKT discrepancies--i.e., item difficulties were different from the tasks they were supposed to represent. When the absolute values of the "sampling," "difficulty," and "error" rows were added, an average of 49.6 percent of the absolute discrepancies were accounted for by the difficulty factor. Sampling and interaction accounted for 17.0 percent and 33.4 percent of the discrepancies, respectively.

In a few duty areas, the error component had a large impact, counteracting otherwise large HOPT-JKT discrepancies. For Form A, Communications, NBC, and Security and Intelligence would have been much more discrepant had it not been for a large amount of error. For Form B, error counteracted discrepancies for Night Vision, and Security and Intelligence.

Table 4. Results of sample-difficulty analyses of JKT differences from the HOPTs

	Sampling Component	Difficulty Component	Error Component	Total Discrepancy
Form A				
Communications	+.19	+.22	-.30	+.11
First aid	+.02	+.22	-.06	+.18
Grenade launcher	+.01	+.07	-.01	+.07
Hand grenades	0.00	+.03	0.00	+.03
LAW	+.02	+.11	-.02	+.11
Land navigation	+.03	+.10	-.10	+.03
Mines	+.04	+.01	+.01	+.06
NBC	+.07	+.30	-.26	+.11
Night vision	+.02	+.23	-.02	+.23
SAW	-.01	+.15	-.12	+.04
Security & Intel.	+.03	+.25	-.17	+.11
Tactical Measures	-.09	+.19	0.00	+.10
Form B				
Communications	+.15	+.12	-.18	+.09
First aid	+.05	+.07	-.08	+.04
Grenade launcher	+.01	+.08	0.00	+.09
Hand grenades	0.00	-.01	0.00	-.01
LAW	+.03	+.11	-.02	+.12
Land navigation	+.13	+.18	-.26	+.05
Mines	+.02	-.13	-.01	-.12
NBC	0.00	+.16	-.17	-.01
Night vision	+.01	+.51	-.33	+.19
SAW	0.00	+.20	-.13	+.07
Security & Intel.	+.30	+.20	-.34	+.16
Tactical Measures	-.09	+.19	0.00	+.10
TOTAL ABSOLUTE	1.32 (17.0%)	3.84 (49.6%)	2.59 (33.4%)	7.75 (100.0%)

NOTE: Positive values indicate that the respective component (i.e., factor) results in the total hands-on average being higher than the total job-knowledge average; negative values indicate that the component makes the HOPT average lower. The sampling component will be positive if the easier HOPT tasks constitute a larger proportion of the HOPT total score than the corresponding JKT items. The difficulty component is positive if HOPT tasks are easier than the average of corresponding items. The error component includes the systematic interaction of sampling with difficulty and random error. It will be large and negative if the biggest difficulty differences were for tasks that also had substantial sampling discrepancies, and the signs of the discrepancies match.

Item-Level Discrepancy Analyses

S-D analyses, described above, compared average item difficulties to HOPT percentage correct *at the duty-area level*. This section presents item-level discrepancy analyses. Since the difficulty component was found to account for the largest proportion of duty-area HOPT-JKT discrepancies, additional analyses were conducted to account for why these occur. To supplement the S-D analyses of average item difficulties just described, single item-level discrepancy analyses were conducted using the crosswalk described above and shown in appendix D.

For each JKT item, the average percentage of steps correct for each HOPT task was subtracted from the corresponding JKT item percentage correct. Positive discrepancy values therefore refer to cases in which the HOPT task was easier than the corresponding JKT item. Negative values indicate cases where the HOPT was more difficult than the job-knowledge item. Appendix E lists the items from lowest to highest discrepancy values.

Figure 3 illustrates the results of plotting task percentage of steps correct (values are labeled along the y-axis) versus the average percentage of examinees responding correctly to the item (values along the x-axis). Tasks lying along the diagonal line were of the same difficulty as the corresponding item. Distance from the diagonal line indicates the degree of item-task discrepancy. Tasks above the line were easier than the corresponding item, and those below the line were more difficult than the matched item.

Figure 3 indicates that there is considerable variation not accounted for by duty areas. For example, although the average discrepancy for the land navigation area is quite low, several items concerning determination of location by map-terrain association and setting an azimuth at night was considerably easier than the performance of corresponding HOPT. Similarly, although JKT performance in the communications duty area was generally worse than the performance of corresponding HOPT tasks, items for one task, constructing a field-expedient antenna, were much easier than the HOPT.

Substantive Analyses of Item-Task Differences

The analyses so far indicate that difficulty is a more important factor than sampling problems in determining HOPT-JKT discrepancies. The following section explores three reasons for these discrepancies in item difficulty:

- The item writing was misleading or the item context was missing.
- The wrong skill is measured by the item.
- The item was written for a skill level different from that possessed by the examinees.

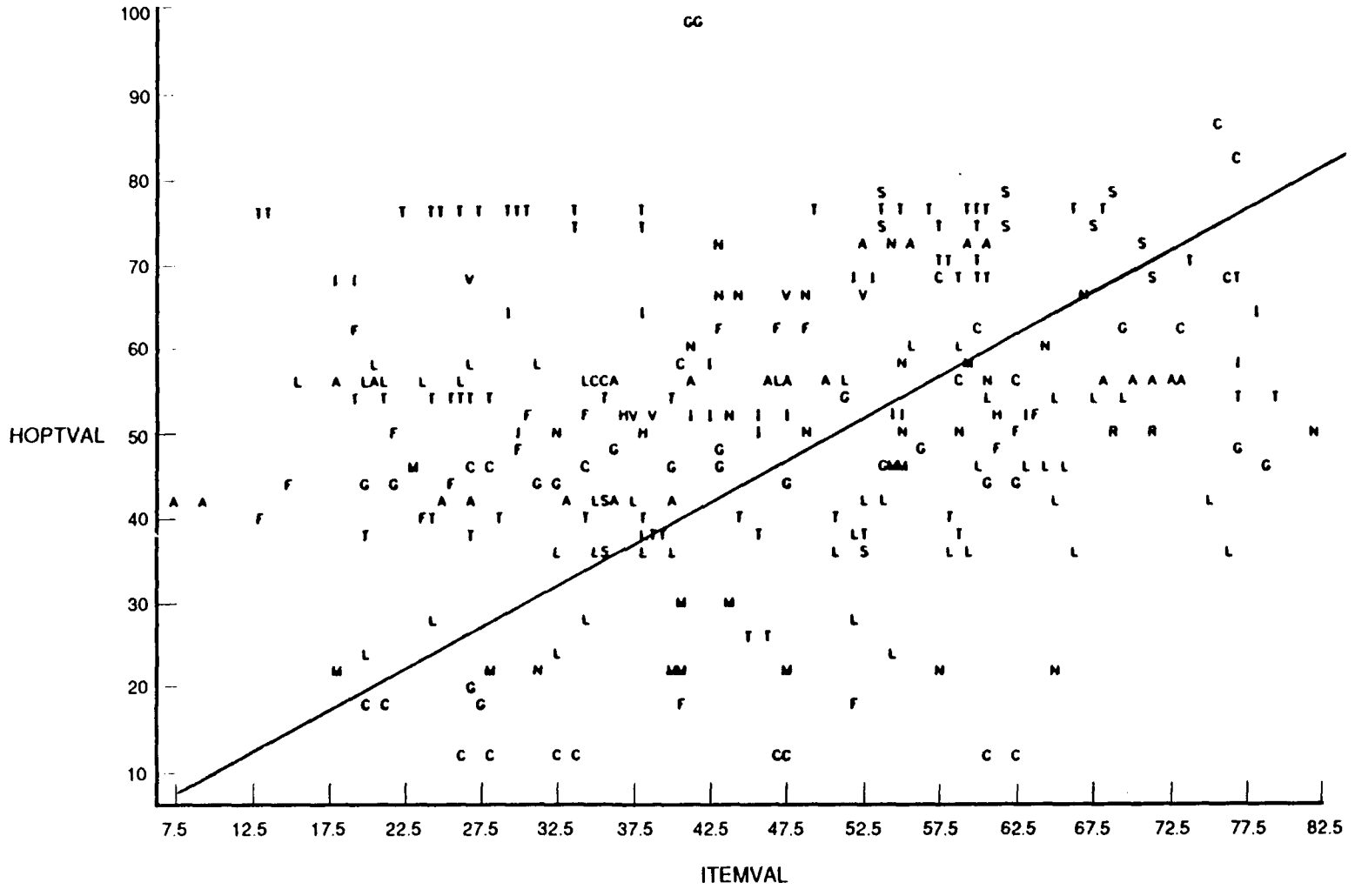


Figure 3. Plot of HOPT percentage correct vs. JKT item percentage correct

Misleading or Unclear Context

It is expected that even if Marines do not know the correct answer, they will guess the correct answer roughly 25 percent of the time for a four-alternative test. If the percentage of correct responses is considerably below the chance level, it is suspected that the item might have been misleading in some way. Appendix F lists items that had less than a chance level of correct responses (25 percent) for either Form A or Form B. Three items stand out as possibly misleading: LAW1, TM17, and FA4.

LAW1, answered correctly 9.4 percent of the time for Form A, asked what you should do *first* if the LAW does not fire. The correct answer was "squeeze the trigger," but the alternative "wait 10 seconds and fire again" was considerably more popular. In fact, waiting 10 seconds and firing again is the *second* thing that should only be done--it should only be performed if resqueezing the trigger does not work. In this case, Marines might have assumed that the trigger had been resqueezed, so the context of the question was unclear. This shows one danger of using a job-knowledge test for measuring all aspects of a job--the paper-and-pencil item asked for the Marine to think about a reflex response. It is inadvisable to use multiple-choice items to assess reflexes.

TM17, answered correctly 13 percent of the time for Form A, was written to reflect the task "Control Unit When Not In Contact." This question asked when the Marine should use an alternate route from an objective, with two of the answer choices being (c) as needed, to avoid contact with the enemy, and (d) when you used the primary route to reach the objective. In this case, the correct answer was "d." This question was somewhat misleading because the purpose of using the alternate route is to avoid contact with the enemy--so "c" was very often chosen. In this sense, the response choices were not mutually exclusive.

A third example was CPR (FA4). This item asked, "When giving CPR, how often should you check for breathing and pulse?" Choices were (a) after each compression/cycle, (b) after every two cycles, (c) after every three cycles and (d) after every four cycles. The correct answer was after every four cycles. In this case, it seems like a fair question, although Marines apparently did not know the answer--it was answered correctly only 13 percent of the time for Form A.

Low item-total correlations also indicate misleading items, because low correlations indicate that those who did best overall were choosing an alternative other than the one on the answer key. Table 5 shows items that had low correlations with total score. It is noteworthy that the proportion of successful responses for these items was nearly always at the chance level--10 of the items were also in the list of items with lowest frequency of correct responses.

Table 5. Items with low item-total correlations

Form A			Form B		
Item	Correlation	Mean Proportion Correct	Item	Correlation	Mean Proportion Correct
FA5	.03	.22	LAW1	.00	.08
GL3	.05	.40	LAW14	-.04	.18
GL9	.03	.31	LN10	-.02	.21
LAW1	.05	.09	MI4B	-.05	.18
LAW14	-.05	.20	MI6B	.05	.28
LN6	.01	.20	TM11	.01	.20
NBC1B	-.07	.20	TM12	.01	.26
TM11	.01	.21	TM14	-.04	.28
TM12	-.04	.26	TM16	.02	.26
TM17	.03	.13	TM22	.02	.27
TM18	03	.28	CM13	.04	.20
TM35	.05	.36	CM14	.01	.35
CM10	.02	.26			
CM13	-.03	.21			
CM14	-.03	.35			

Table 6 shows that when low-correlation items were deleted, the correspondence between the JKT and the HOPT increased. For Form A, the correspondence increased more than 44 percent for those duty areas that were changed, and 20 percent overall. For Form B, the improvement was 27 percent for duty areas that were changed, and 14 percent overall.

Items representing the ability to work on mines with tripwires (Form B) were exceptions to the rule that deleting low-correlation items improved HOPT-JKT correspondence. This was apparently because, in fact, Marines had little proficiency in this hands-on skill. This anomaly indicates that multiple-choice items are inadequate for measuring proficiency in performing extremely difficult tasks, because random guessing provides a lower bound on the proportion of correct responses. In all other instances, deleting low-correlation items increased the correspondence between JKTs and HOPTs.

Table 6. HOPT–JKT discrepancy values after low item—total correlation items are omitted

	Original no. items/ no. deleted	Sampling component		Difficulty component		Error component		Total discrepancy		Total percentage improvement with deletions
		Before deletions	After deletions	Before deletions	After deletions	Before deletions	After deletions	Before deletions	After deletions	
Form A										
Communications	15/3	+.19	+.16	+.22	+.23	−.30	−.33	+.11	+.06	+45%
First aid	8/1	+.02	+.02	+.22	+.26	−.06	−.12	+.18	+.16	+11%
Grenade launcher	12/1	+.01	−.01	+.07	+.05	−.01	0.00	+.07	+.04	+43%
LAW	13/2	+.02	+.02	+.11	+.08	−.02	−.03	+.11	+.07	+36%
Land navigation	24/2	+.03	+.02	+.10	+.09	−.10	−.11	+.03	0.00	+100%
NBC	11/1	+.07	+.07	+.30	+.28	−.26	−.27	+.11	+.08	+27%
Tactical measures	35/8	−.09	−.08	+.19	+.16	0.00	−.03	+.10	+.05	+50%
Form B										
Communications	13/3	+.15	+.11	+.12	+.13	−.18	−.20	+.09	+.04	+56%
First aid	10/1	+.05	+.05	+.07	+.07	−.08	−.08	+.04	+.04	0%
Grenade launcher	12/2	+.01	−.01	+.08	+.06	0.00	+.01	+.09	+.06	+33%
LAW	13/2	+.03	+.02	+.11	+.07	−.02	−.03	+.12	+.06	+50%
Land navigation	24/2	+.13	+.13	+.18	+.17	−.26	−.28	+.05	+.02	+60%
Mines	7/2	+.02	+.01	−.13	−.18	−.01	−.01	−.12	−.18	−50%
Tactical measures	35/8	−.09	−.08	+.19	+.17	0.00	−.03	+.10	+.06	+40%

NOTE: Positive values indicate that the component results in the job–knowledge test being more difficult than the hands–on performance test. "Total percentage improvement with deletions" was computed using the original total discrepancy as a baseline. For example, the total percentage improvement with deletions for Form A, Communications, was computed .11−.06/.11=45%.

The Wrong Skill Level Is Measured

Measuring a different skill level was a common reason for an item to be discrepant from the HOPT. For example, items measuring complex procedural tasks were often too easy. For constructing a field-expedient antenna, the HOPT mean was 12 percent, but the item means were generally much higher. The actual HOPT task was complex (appendix F), requiring the ability to read detailed directions, visualize performance steps, organize multiple pieces of equipment, and demonstrate a moderate amount of manual dexterity. There are a number of ways performance steps could be misinterpreted, even with help from the manual. (All Marines were allowed to use manuals to assist with this task). In contrast, the corresponding item, CM9 (mean 62.8 percent), required no more than the ability to read a chart.

CM9, meant to measure the ability to construct a field-expedient antenna, could be improved if it focused on a more appropriate step. This item focused on a step that was probably one of the least difficult in the entire task. Furthermore, an approach using fill-ins for "what would you do next" would come closer to matching the difficulty of the actual performance test.

Items measuring skill at complex reasoning tasks were also often too easy. For example, setting an azimuth at night requires the ability to follow complex instructions, visualize, and make independent judgments. On the other hand, a corresponding item stressed simple memory of definition of a back azimuth, resulting in a comparatively high pass rate (54.1 percent and 51.9 percent. It is apparently much more difficult to set an azimuth at night than to remember the definition of a back azimuth.

Better task analysis could have lessened the problems involved with measuring complex procedural and reasoning tasks. In these cases, distractors might have pinpointed common misunderstandings that can interfere with hands-on performance.

When the wrong skill is measured by an item, the result could also be that the JKT is too difficult. A striking example of this was for night-vision items, which were much more difficult than the corresponding HOPT tasks. Several night-vision tasks were relatively straightforward and procedural, resulting in a moderate passing rate of 52 percent for the "clean components" task. In contrast, the corresponding JKT items were overly detailed and asked for relatively unimportant information that a Marine was unlikely to remember. A night-vision "clean components" item on the job-knowledge test, with a 26 percent passing rate, illustrates these more difficult questions as in the following example:

Item 1: What should you use to clean the rubber eyeshield on the AN/PVS-4 night sight?

A. Wet cloth

B. Lint-free cloth

C. Alcohol

D. Soft brush.

In this case, the item asks the Marine to make distinctions among cleaning tools that could be looked up in the manual. The item makes the task too difficult by asking for information that is not needed for successful task performance.

Items measuring time-critical tasks were sometimes too difficult. As described above, the LAW item asking Marines to think about a reflex action was too difficult. Tactical measures was another area in which JKT items were often too difficult compared to HOPT performance. TM item 17 was difficult (item mean 13 percent, task mean 76.8 percent). This item asks for information about tactics that a first-term Marine would not be required to know. In addition, the item is not good as a surrogate because it measures reading comprehension and carefulness: the preposition "from" in the item stem ("from an objective") is crucial to understanding that the correct answer is D. This item is therefore a poor surrogate because, to a large extent, it tests reading comprehension rather than understanding of tactical measures. Furthermore, someone could have had partial understanding that is not given credit in the following "all or none" example:

Item 17: You have completed a detailed patrol plan, and selected an alternate route from an objective. You should use that alternate route

A. only when the patrol made contact with the enemy on the primary route

B. only when the patrol leader suspects that the patrol has been detected

C. as needed, to avoid contact with the enemy

D. when you used the primary route to reach the objective.

A Non-Matching Skill Level Is Measured

Many tactical measures (TM) items measured skill levels that were more advanced than those yet acquired by most examinees.[1] The TM tasks often involved knowing instructions (e.g., hand signals for various formations) that first-term Marines would know, but items focused more on knowing the conditions under which certain formations should be used --which is material unfamiliar to first-term Marines.

TM item 34 (Control Unit Movement When Not In Contact) also measured a non-matching skill level. The item used a system of markings that are unfamiliar to most first-term infantrymen (figure 4). The HOPT task (appendix H) gives the infantryman more information about the tactical situation, showing the situation on a map, whereas the item gives only the ambiguous information "you want to advance rapidly across a danger area against a known enemy position." The HOPT task is comparatively easy because it asks a series of questions familiar to what first-term Marines should know (e.g., proper hand signals for various formations), whereas the item focuses on knowledge that would be available mostly to squad leaders (e.g., which formation to use). Note that the HOPT task allows for partial knowledge (by scoring some steps correctly and others incorrectly), whereas the JKT item is "all or none." Lastly, note that the JKT item does not make clear in which direction the squad leader should move his troops, therefore, important context is missing.

CONCLUSIONS

Implications for Development of Better Job-Knowledge Tests

Better Control of Item Sampling

An immediate implication of this research is that *job-knowledge tests should more adequately sample the domain, sometimes with more items.* This research showed that poor item sampling or the interaction of sampling with difficulty was responsible for about 50 percent of the discrepancies between JKT and HOPT duty-area totals. To the extent possible, the proportion of items for a duty area should reflect the proportion of tasks performed in the duty area. Otherwise, JKT and HOPT averages might be discrepant solely because the balance of JKT items reflects tasks that are unrepresentative of the duty area as a whole.

---

1. Some JKT items and HOPT tasks were purposefully developed to deal with content relatively unfamiliar to most first-term Marines. This was done to determine the ability of Marines to respond if unfamiliar leadership roles were thrust upon them during combat. Although these items and tasks were purposefully developed, they sometimes increased HOPT-JKT discrepancies.

You want to advance rapidly across a danger area, against a known enemy position. As squad leader, which squad combat formation should you set up?

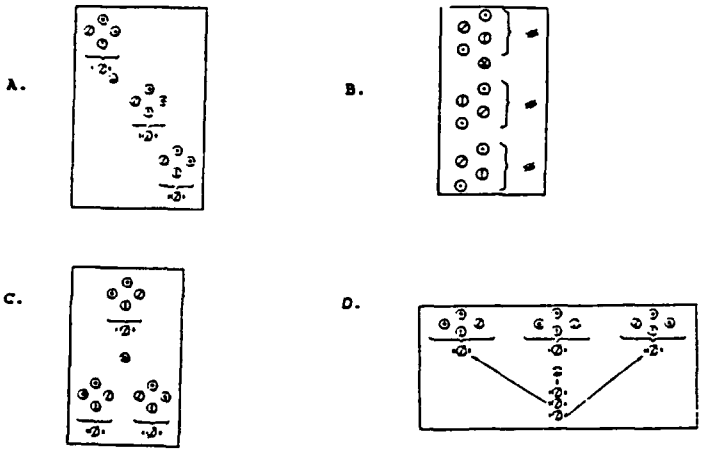


Figure 4. Tactical Measures Item 34, Control Unit Movement When Not in Contact

To correct the problem of poor item sampling, content domain specifications need to be made more explicitly, taking care to give approximately the same proportion of items as there are tasks in the content domain. Sometimes this will require that a larger number of items be written. It would be relatively simple to keep track of the optimum number of items per task given information on number of tasks and constraints on the total number of items allowed. For example, Communications for Form A had the following proportions of tasks and items:

	HOPT Tasks		JKT Items		Discrepancy
	No.	Fraction	No.	Fraction	
Operations inspection	1	(.1111)	2	(.1429)	-.0318
Visual inspection	1	(.1111)	1	(.0714)	.0397
Operate AN/PRC-77	1	(.1111)	0	(.0000)	.1111
Assemble radio AN/PRC-77	1	(.1111)	1	(.0714)	.0397
Take immediate action	0	(.0000)	1	(.0714)	-,9614
Construct field expedient antenna	1	(.1111)	4	(.2857)	-.1746
Install telephone set	1	(.1111)	2	(.1429)	-.0318
Repair wire of TA-312	1	(.1111)	1	(.0714)	.0397
Operate TA-312	1	(.1111)	0	(.1429)	.1318
Check parts	1	(.1111)	0	(.0000)	.1111
Total	9	1.0000	14	(1.0000)	

If the 14 items had been distributed so that "Construct Field Antenna" had fewer items, and at least one item had been written for each task, sampling would have improved considerably. The discrepancy would have decreased further if the item measuring "Take Immediate Action" had been deleted, since there was no corresponding task.

Quality Control Procedures for Items

Another important step in creating a better job-knowledge test involves more complete integration of task analyses with item development. Item writers should think about which steps are likely to be performed incorrectly for lack of knowledge, and write items for those steps. Ideally, distractors would include common errors made by Marines. Item writers should write reasoning items for reasoning tasks, and not simplify an unstructured task into a simple choice of alternatives (as was done for setting an azimuth at night and determining grid coordinates). Item writers should avoid contextual ambiguity in their items and should be careful not to write items that stress mainly reading ability, or which involve overly detailed knowledge that the Marine is likely to look up in a manual (e.g., cleaning night-vision equipment). Appendix I provides a group of questions that could be used for item review.

Try-out testing is another important part of improving item quality. More care needs to be taken to delete items that have low item-total correlations. Improvements can be dramatic if such items are deleted. To delete items that have low item-total correlations, it would be necessary to "trial test" the job- knowledge test, compute the correlations, delete the low correlation items, and "trial test" the remaining items a second time. Results from the second trial test should confirm that all poor items have been eliminated.

Alternate Formats for Job-Knowledge Tests

For skills that will be hard to measure with a traditional multiple-choice job-knowledge test, a variety of alternative formats are available (table 7). Haladyna [11] suggests these alternatives because they can require complex thinking without introducing irrelevant response alternatives. Some research has shown that, in practical test writing, only one or two distractors carry most of the burden. Alternate choice (table 7) is useful for avoiding overly-detailed distractors and decreasing the reading ability needed to answer the question. Multiple true-false and testlets, also shown in table 7, can simulate some of the complexity of lengthy procedures. Fill-in and essay questions are also alternatives to traditional multiple choice. If a job-knowledge item is intended to measure ability to perform difficult procedural or reasoning tasks, use of question formats such as "What is the correct way to...?," "Which is the most important?" and "What would happen if...?" might also improve item validity (Table 8, from Haladyna, 11).

Table 7. Examples of nontraditional job-knowledge test item formats

---

Alternate Choice

You have completed a detailed patrol plan and selected an alternate route from an objective. Under which condition should you use that alternate route?

1. To avoid contact with the enemy

2. When you used the primary route to reach the objective.

Multiple True-False

You have completed a detailed patrol plan and selected an alternate route from an objective. Which of the following are conditions under which you should use that alternate route?

1. When the patrol made contact with the enemy on the primary route

2. When the patrol leader suspects that the patrol has been detected

3. To avoid contact with the enemy

4. When you used the primary route to reach the objective.

Item Set (Testlet)

Suppose you have to set an azimuth of 45 degrees at night. You have rotated the bezel ring until the luminous line is directly over the black index line.

1. What should now be rotated?
   a. Bezel ring b. yourself

2. In what direction should rotation be accomplished?
   a. clockwise b. counter-clockwise

3. How much rotation is needed?
   a. 9 clicks b. 15 clicks

Table 8. Examples of generic item shells for complex tasks

Applying	Predicting
What is the correct way to.... Background is given. o What is the problem? o What is the solution to the problem? o How should the problem be solved?	What would happen if...? When..., what happens? Under what circumstances would you expect....?

IMPLICATIONS FOR WHEN A JKT IS APPROPRIATE

If attempts with nontraditional item formats are unsatisfactory, alternatives to a job-knowledge test should be sought. Physical coordination tasks such as firing a rifle and reflexive, time-critical tasks such as taking immediate action are very difficult to measure validly with a JKT. Analysis of the 50 largest item-task discrepancies from appendix E suggests that complex reasoning tasks, such as setting an azimuth at night, and highly procedural tasks, such as constructing a field-expedient antenna, are also difficult to measure with a job-knowledge test. In summary, the following types of tasks might be inappropriate to test using a paper-and-pencil measure:

- *Physical coordination tasks*, such as firing a rifle at pop-up targets are impossible to measure with a job-knowledge test [5].
- *Time-critical, reflex tasks*, such as responding to a LAW that does not fire might be impossible to measure with a job-knowledge test.
- *Complex procedural tasks* that require multiple steps and the ability to visualize the completed project, such as construction of a field-expedient antenna, were also difficult to test using multiple-choice tests.
- *Complex reasoning tasks* that require spatial orientation, such as setting an azimuth at night, were difficult to test using multiple-choice tests.

In contrast, the following tasks require skills for which testing using a JKT is highly appropriate:

- *Knowledge-driven tasks* that require memory for specific facts and attention to detail among complex alternatives. These tasks require knowledge, but it is helpful not to have to look the information up in a textbook.

- *Reading-dependent tasks* in which the JKT is a close approximation to the form in which actual job performance is required. Skills in using a technical manual are especially appropriate for a JKT.

- *Time-independent tasks* in which the actual job performance allows sufficient time to recall information; some simple maintenance tasks might be fairly time-independent.

REFERENCES

[1] Committee on the Performance of Military Personnel, Commission on Behavioral and Social Sciences and Education, National Research Council, *Job Performance Measurement in the Military: Report of a Workshop*, Washington, DC: National Academy Press, 1984

[2] CNA Research Memorandum 90-47, *An Assessment of Surrogates for Hands-On Tests: Selection Standards and Training Needs*, by Neil B. Carey, Jul 1990. (27900047)[1]

[3] CNA Research Memorandum 89-311, *Methods to Assess the Utility of Proxies*, by Neil B. Carey, Aug 1990 (27890311)

[4] Carey, Neil B., *Surrogates for HOPTs: Selection Standards and Training Needs*, Paper presented at the 98th Annual Convention of the American Psychological Association Symposium entitled "How Good are Alternative Measures of Job Performance?" Boston, MA: Aug. 10-14, 1990

[5] CNA Research Memorandum 90-55, *Alternative Measures of Rifle Firing Skills*, by Neil B. Carey, Sep 1990 (27900055)

[6] CNA Research Memorandum 90-57, *A Shift-Share Analysis of First-Term Attrition*, by Timothy W. Cooke, with James E. Grogan and Roberta T. Taggart, Dec 1990 (27900057)

[7] CNA Research Memorandum 89-45, *Performance of Basic Infantry Tasks*, by Paul W. Mayberry, Jul 1989 (27890045)

[8] CNA Research Contribution 570, *Developing a Competency Scale for Hands-On Measures of Job Performance,* by Paul W. Mayberry, Unclassified, Dec 1987 (27890045)

[9] Richard J. Shavelson et al. "Generalizability of Job Performance Measurements: Marine Corps Rifleman." *Military Psychology*, 2(3), 1990: 129-144

[10] CNA Research Memorandum 90-110, *Incremental Validity of New Tests in Prediction of Infantry Performance*, by Paul W. Mayberry and Catherine M. Hiatt, Aug 1990 (27900110)

1. The number in parentheses is a CNA internal control number.

REFERENCES (Continued)

[11] Haladyna, T.M. "Test Item Design and Evaluation," Paper presented to the Invited Symposium of the International Personnel Managers Association, San Diego, California, June 25, 1990

[12] Hambleton, Ronald K. "Test Score Validity and Standard-Setting Methods," in *Criterion-Referenced Measurement: The State of the Art*, pp. 80-123. Baltimore, MD: The Johns Hopkins University Press, 1980

APPENDIX A:

DETAIL OF HOPT DEVELOPMENT PROCEDURES AND RELIABILITY

APPENDIX A

DETAIL OF HOPT DEVELOPMENT PROCEDURES AND RELIABILITY

Development

The first task in developing job-performance measures was to define the requirements of Marine Corps enlisted infantrymen for each Military Occupational Specialty (MOS). This was essential to ensure that tasks would be selected to maximize the coverage of job behaviors so that hands-on test scores would generalize to the full range of infantry job requirements.

Extensive job analyses using Marine Corps manuals on the performance of job tasks were conducted to define hands-on performance content. Matrices of task and skill requirements for each MOS were developed and reviewed by Marine Corps job experts. Test content was randomly selected from these matrices so that scores could be generalized to the full performance domain.

Job tasks were organized into duty areas, and each duty area was covered by one or more performance tests. The number of tasks performed for each of the basic infantryman duty areas varied between one (for hand grenades) to nine (for land navigation).

Second, job experts identified specific skills and knowledge required to perform the job of each MOS. Thus, underlying skills and knowledge common across MOSs were made explicit. This procedure allowed common and unique skills and knowledge to be sampled across MOSs.

Hands-on performance tests (HOPTs) were developed for the selected test content. These HOPTs were reviewed by Marine Corps job experts. The tests were then trial-tested and improved before a large-scale tryout was conducted with more than 200 Marines. Test tryout and test-administrator training were conducted during the first two weeks of August 1987. A full two-week training period for test administrators was conducted because of the critical nature of the scorers' grading judgments.

Detailed task analyses of the selected test content were then conducted to identify the specific steps required to perform each task. Job experts and job incumbents reviewed the task analyses to confirm their validity as accurate descriptions of how tasks were actually performed.

Each task required the infantryman to perform a series of steps that would be scored either "go" or "no-go." Some tasks had as few as 2 steps and others as many of 37, but most tasks contained approximately 10 steps.

Two testing forms were developed for each MOS. The number of tasks given each participant ranged from 68 to 71 (for riflemen), 70 to 72 (for machine gunners), 72 to 75 (for mortarmen), and 76 to 80 (for assaultmen).

Trial tests of representative tasks for the rifleman specialty (MOS 0311) were administered to more than 200 Marines, to ensure that tasks could be completed and scored under actual test conditions. The tryout was also used to train the test administrators to achieve and maintain equivalent scoring standards across testing situations. Tryout of the tests immediately followed administrator training.

The most critical component of hands-on performance measurement is the test administrator. Unlike paper-and-pencil tests in which reliable and objective scoring keys are easily applied, hands-on administrators must observe and make judgments concerning whether individuals performed each step correctly. Former Marines were hired to serve as test administrators because of their familiarity with the test content, knowledge of the Marine Corps, and ability to work well with young Marines. Because test administrators were retired, they did not have a vested interest in coaching or scoring some Marines more leniently.

To ensure comparability of hands-on scoring across testing locations, detailed training manuals were prepared, and the same testing team conducted the training at each base. To monitor the scoring accuracy and consistency of the test administrators, daily quality control checks were implemented. Hands-on data were entered into a computer daily so that administrators could be checked for leniency or drift. Immediate, specific feedback was given to test administrators if problems were detected. To assess the accuracy of hands-on scoring, shadow scoring was conducted as a quality control on a regular basis. Discrepancies among administrators' scoring were discussed and resolved. Administrators were rotated across testing stations to minimize systematic error and to increase administrator motivation and attention.

Reliability

Generalizability analyses [9] indicate that the HOPT used in this research sampled enough tasks to have a relatively high G coefficient of .83 using data from only 35 of the more than 70 tasks used to make up the HOPT. The major sources of variation with these data concern tasks; almost negligible error is associated with examiners [9]. These analyses ensure that the procedures used to develop the HOPT were successful in creating a test that can be generalized to the full domain of infantry performance.

Interrater agreement, measured in percentages, is the number of times raters agree on their markings, divided by the total number of steps marked. Scorer agreement ranged from a low of 80 percent to a high of 100 percent, depending on the task. The mean interrater

agreement was 90 percent. Agreement levels for the four MOSs were 0.90, 0.90, 0.89, and 0.90 for MOS 0311, 0331, 0341, and 0351, respectively. These agreement levels compare favorably with other studies of hands-on performance.

Test-retest reliability is the degree to which people score the same on a subsequent test administration. The reliability of the hands-on measures was tested with 188 riflemen (MOS 0311) taking the opposite form of the test seven to ten days after the initial administration. For example, a rifleman who took the 68-item Form A test originally would take the 70-item Form B test seven to ten days later. The correlation (reliability estimate) between the two administrations was 0.70. Significant carryover effects were found: there was an average retest gain in performance of more than 0.8 standard deviation.

Internal consistency is the degree to which different items on a test indicate the same level of proficiency. Cronbach alpha coefficients, which are estimates of internal consistency, were computed for each MOS and each test form. Alpha coefficients were 0.87 for MOS 0311 (n = 1,067), 0.87 for MOS 0331 (n = 257), 0.88 for MOS 0341 (n = 217), and 0.83 for MOS 0351 (n = 239). In no cases did alpha coefficients for alternate forms vary by more than 0.02. These figures indicate a high and stable degree of internal consistency for the hands-on tests.

Table A-1. Reliability estimates of hands-on test

		Reliability		
		Test-Retest	Internal Consistency	Scorer Agreement
Hands-on	Form A	0.70	0.88 (69 tasks)	0.90
	Form B		0.86 (66 tasks)	0.90

APPENDIX B:

DETAILS OF JKT DEVELOPMENT PROCEDURES AND RELIABILITY

Appendix B

DETAILS OF JKT DEVELOPMENT PROCEDURES AND RELIABILITY

Development

The paper-and-pencil job-knowledge tests were developed to parallel hands-on content as much as possible. Therefore, hands-on performance steps were used as a basis for developing the job-knowledge test, although it was understood that the time allowed for hands-on testing would not permit all hands-on content to be covered.

Beginning with the procedures specified in the hands-on performance test, critical steps were identified and multiple-choice questions were written concerning those steps. The items stressed what and how steps are performed rather than why. Whenever possible, illustrations were used to maximize the fidelity to actual performance situations. For tasks that were more cognitive than procedural (e.g., tactical measures), combat scenarios were developed, and items asked what should be done based on the information provided in the scenarios.

Items were reviewed by Marine Corps subject experts, then by test-development experts, for psychometric qualities. Written test forms were developed corresponding to the content for Form A and Form B of the general infantry hands-on test items. Content that was the same in both hands-on test forms corresponded to the same written items. Two sets of prospective items for each test form were developed and labeled A1, A2, B1, B2. The order of duty-area presentation was varied between each version of Form A or Form B, but the order remained the same within each duty-area. The initial draft of the job-knowledge test contained 175 items, with the expectation that information from the tryout would eliminate some items. Only one written test form was developed to parallel each of the MOS-specific parts of the hands-on tests.

The draft job-knowledge test questions were evaluated using active-duty Marines. Seventy-one Marines completed 175 items for the general infantry questions and 42 specific Rifleman (0311) items. Two alternative versions of the general infantry test were administered. Completion times were recorded, and item analyses pinpointed items that were keyed incorrectly and items that should be deleted. Items were dropped that had high correlations of distractors with total test score or an abnormally low pass rate. Either of these conditions indicates that the item was keyed incorrectly or ambiguous enough that the more knowledgeable Marines did not do better on the item.

The final general infantry (0300) test consisted of 150 items, to be completed in 90 minutes. The number of MOS-specific items varied from 40 to 50.

Reliability

Test-retest reliability, which measures the correlation between two administrations of the same job-knowledge test administered seven to ten days apart to 189 riflemen, was 0.73 [10]. This degree of reliability is adequate, but not particularly high.

Internal consistency, estimated by Cronbach alpha coefficients, computed for the job-knowledge test, was 0.89 for MOS 0311 (199 items, $\underline{n}$ = 1,296), 0.89 for MOS 0331 (190 items, $\underline{n}$ = 306), 0.90 for MOS 0341 (189 items, $\underline{n}$ = 312), and 0.87 for MOS 0351 (190 items, $\underline{n}$ = 314). The difference between alternate test forms never varied by more than 0.02 for any MOS. These figures indicate that different parts of the job-knowledge test were measuring the same skills.

APPENDIX C:

DESCRIPTION OF SAMPLE-DIFFICULTY ANALYSES

APPENDIX C

DESCRIPTION OF SAMPLE-DIFFICULTY ANALYSES[1]

Sample-difficulty analysis allows decomposition of the change in HOPT-JKT discrepancies in observed percentage-correct scores by use of the following algebra: Let i = 1,...,n denote the task, and j=1,2 denote the testing mode (j=1 for HOPT, j=2 for JKT). The percentage of correct steps for a given HOPT task will therefore be denoted r_{i1} and the percentage of examinees passing the set of JKT items representing that task will be denoted r_{i2}. Suppose that there are "n" tasks for the duty area 1,...,n, with the percentage of tasks within the duty area represented by q_{11}, ..., q_{n1}. Each task has an observed percentage of HOPT steps correct r_{11}, ..., r_{n1}. In this case, the percentage of HOPT steps performed correctly for the duty area, P_{HOPT}, can be expressed as

$$\sum_{i=1}^{n} r_{i1}q_{i1} = P_{HOPT} \tag{1}$$

The percentage of items answered correctly is similarly computed as the weighted average of percentage answering correctly to those items that represent a given task for the "n" tasks. The percentage of JKT items correctly answered for the duty area, P_{JKT}, can be expressed as

$$\sum_{i=1}^{n} r_{i2}q_{i2} = P_{JKT} \tag{2}$$

The difference between the percentage of steps performed correctly within the duty area and the percentage of items answered correctly is $P_{HOPT} - P_{JKT}$, which can be expressed as

$$P_{HOPT} - P_{JKT} = \sum_{i=1}^{n} r_{i1}(q_{i1}-q_{i2}) + \sum_{i=1}^{n} q_{i1}(r_{i1}- r_{i2}) - \sum_{i=1}^{n} (r_{i1} - r_{i2})(q_{i1} - q_{i2}) \tag{3}$$

The first term of this expression is a measure of the change in percentage correct that would have been expected if the same percentage of items were answered correctly as task steps (r_{i1} unchanged), but the proportion of items was different than the proportion of tasks ($q_{i1} - q_{i2}$). Calculation of this "sampling" term yields the expected discrepancy in duty-area average because of differences in proportion of

1. Based on work of Cook [6].

items compared to tasks. The remaining two terms are associated with different difficulty levels of tasks and items ($r_{i1} - r_{i2}$). The second term represents "pure" difficulty differences, whereas the third term indicates an interaction of difficulty with sampling, plus error. This third term is called the error component. A table showing an example of data layout and computations is provided in table C-1.

Table C-1. Example of data layout for sample-difficulty analyses (night-vision duty area, Form A)

	HOPT		JKT	
	Average percentage of steps performed (r_{i1})	Percentage of tasks (q_{i1})	Percentage answering relevant items correctly (r_{i2})	Percentage of items (q_{i2})
Task Name				
Visual inspection	.679	.3333	.266	.2500
Operations inspection	.653	.3333	.524	.2500
Clean components	.519	.3333	.382	.5000

NOTE: For the night-vision duty area, there were three tasks ($n = 3$) and four items. There was one item each for visual inspection and operations inspection. There were two items for clean components.

The sampling component for the night-vision duty area is computed as

$$r_{i1}(q_{i1} - q_{i2}) = .679(.3333 - .2500) + .653(.3333 - .2500) + .519(.3333 - .5000)$$

$$= .02 \quad \text{(when rounded to two figures)}$$

The difficulty component for night vision is computed as

$$q_{i1}(r_{i1} - r_{i2}) = .3333(.679 - .266) + .3333(.653 - .524)$$
$$+ .3333(.519 - .382)$$
$$= .23 \text{ (when rounded to two figures)}$$

Note that all of the terms are positive, indicating that all shifts result in the job-knowledge items being more difficult than the corresponding HOPT tasks. Therefore, the sign is positive.

The final "error" term, is computed as

$$(r_{i1} - r_{i2})(q_{i1} - q_{i2}) = (.679-.266)(.3333-.2500)$$
$$+(.653-.524)(.3333-.2500)$$
$$+(.519-.382)(.3333-.5000)$$
$$= .02 \text{ (when rounded to two figures)}$$

Note that this term is subtracted from the other two in equation (3). A large error term would have negated the main effects for sampling or difficulty.

Taken as a whole, the three terms indicate that the difference between task and item difficulties, rather than sampling or interaction, is the primary reason that the JKT average scores were lower than for the HOPT.

APPENDIX D:

HOPT TASK AND JKT ITEM DIFFICULTIES COMPARED

APPENDIX D

HOPT Task and JKT Item Difficulties Compared

Duty Area	HOPT Tasks	Steps	HOPT Mean	JKT Items	JKT Mean
Land Navigation					
	Setting azimuth *during night*	3	A: 24.9 (*LN01A*)	LN06	20.3
				LN08	32.8
				LN07	54.1
			B: 27.4 (LN01B)	LN06	24.1
				LN08	34.4
				LN07	51.9
	Pace distance (LN02AB)	5	A—59.7	LN09	A—58.7
			B—59.0	LN09	B—55.9
	Determine location by map-terrain association (LN03AB)	1	A—41.9	LN17	75.3
				LN18	53.8
				LN19	35.1
				LN20	34.9
				LN21	64.7
				LN22	52.7
				LN23	37.5
			B—36.9	LN17	76.1
				LN18	59.2
				LN19	32.4
				LN20	38.1
				LN21	66.3
				LN22	58.0
				LN23	40.0
	Determine azimuth *from one point* to another (LN04AB)	3	A—55.3	LN24	19.8
			B—55.5	LN24	15.7
	Convert azimuth	2	A—51.2 (*LN05A*)	(none)	——
			B—50.0 (LN05B)	(none)	——
	Determine grid coordinates	2	A—45.7 (*LN06A*	LN13	65.5
				LN14	63.0
				LN15	59.7
				LN16	64.2
			B—54.4 (LN06B)	LN13	69.5
				LN14	64.7
				LN15	60.6
				LN16	67.4
	Determine location by resection	4	A—38.7 (*LN07A*)	LN01	51.8
				LN05	38.3
			B—36.9 (LN07B)	LN01	50.9
				LN05	35.0

Duty Area	HOPT Tasks	Steps	Mean	JKT Items	Mean
Land Navigation (con't)					
	Determine location by intersection	3	A—56.2 (LN08A)	LN02	23.9
				LN03	51.5
				LN04	46.7
			B—not tested	LN02	27.1
				LN03	53.3
				LN04	45.8
	Follow azimuth	1	A—not tested	(none)	——
			B—82.8 (LN09B)	(none)	——
	Measure distance on a map (LN11AB)	6	A—58.2	LN10	20.5
				LN11	31.2
				LN12	26.6
			B—56.3	LN10	21.0
				LN11	34.3
				LN12	26.0

Duty Area	HOPT Tasks	Steps	Mean	JKT Items	Mean
Tactical Measures					
	Move individually (TL01AB)	12	A: 70.1	TM04	73.7
				TM07	57.9
				TM08	59.9
				TM09	57.7
			B: 68.1	TM04	76.6
				TM07	60.2
				TM08	60.5
				TM09	58.5
	1-Man carries (TL03AB)	9	A: 53.4	(none)	——
			B: 55.3	(none)	——
	Estimate range (TL04AB)	5	A: 27.7	(none)	——
			B: 28.5	(none)	——
	Camouflage self and equipment (TL05AB)	9	A: 74.4	TM01	57.4
				TM02	38.2
				TM03	59.8
				TM05	57.3
				TM06	33.5
			B: 76.0	TM01	60.9
				TM02	37.9
				TM03	59.5
				TM05	53.9
				TM06	33.7
	Establish landing zone (TM01AB)	5	A: 53.3	TM10	79.6
				TM11	21.3
				TM12	25.5
				TM13	24.4
				TM14	27.1
				TM35	35.9
			B: 53.5	TM10	76.8
				TM11	19.6
				TM12	26.3
				TM13	26.0
				TM14	28.3
				TM35	40.0

	HOPT			JKT	
Duty Area	Tasks	Steps	Mean	Items	Mean
Tactical Measures (con't)					
	Direct helicopter landing and takeoff (TM08AB)	6	A: 26.9	TM23	46.2
			B: 25.7	TM23	45.1
	Control unit movement when not in contact (TM09AB)	14	A: 76.8	TM15	68.3
				TM16	30.7
				TM17	13.0
				TM18	27.6
				TM19	55.0
				TM20	29.8
				TM21	25.2
				TM22	27.5
				TM30	57.0
				TM34	24.1
			B: 76.7	TM15	66.1
				TM16	26.2
				TM17	13.5
				TM18	29.2
				TM19	49.4
				TM20	30.6
				TM21	24.2
				TM22	27.4
				TM30	59.8
				TM34	22.8
	Call for and adjust indirect fire (TM14AB)	6	A: 38.1	TM25	45.5
				TM27	39.2
				TM28	59.0
				TM29	52.3
				TM31	38.8
				TM32	20.1
				TM33	26.8
			B: 39.6	TM25	44.4
				TM27	34.1
				TM28	58.4
				TM29	50.5
				TM31	37.9
				TM32	24.1
				TM33	28.6

Duty Area	HOPT Tasks	Steps	Mean	JKT Items	Mean
Night Vision	Visual inspection	1	A: 67.9 (NV01A) B: no task	NV01A (none)	26.6 ——
	Operations inspection (NV02AB)	12	A: 65.3 B: 65.6	NV02 NV02	52.4 47.3
	Clean components	4	A: 51.9 (NV03A) B: no task	NV03A NV04A (none)	37.7 38.7 ——
	Observe	2	A: no task B: 87.2 (NV04B)	(none) (none)	—— ——
	Collect	7	A: no task B: 47.6 (NV05B)	(none) (none)	—— ——

Duty Area	HOPT Tasks	Steps	Mean	JKT Items	Mean
Communications					
	Operations inspection of AN/PRC-77	9	A: 62.5 (CR01A)	CM1 CM4	59.7 57.2
		9	B: 68.2 (CR01B)	CM1 (no others)	57.5
	Visual inspection of AN/PRC-77	4	A: 57.2 (CR02A)	CM5	40.7
			B: no task	(none)	——
	Operate AN/PRC-77	9	A: 84.3 (CR03A)		
			B: no task		
	Assemble radio AN/PRC-77 (CR04AB)	16	A: 82.9	CM2	77.0
			B: 85.9	CM2	75.4
	Take immediate action	9	A: no task	CM3	73.0
			B: 67.4 (CR05B)	CM3	76.0
	Construct field expedient antenna (CR12AB)	15	A: 12.6	CM9	62.8
				CM10	26.1
				CM11	47.7
				CM12	32.6
			B: 12.3	CM9	60.8
				CM10	28.4
				CM11	46.9
				CM12	33.6
	Install telephone set TA-312 (CT07AB)	8	A: 45.9	CM8	26.6
				CM15	34.3
			B: 45.9	CM8	28.1
				CM15	34.6
	Repair wire of TA-312 (CT08AB)	14	A: 18.4	CM13	21.1
			B: 18.9	CM13	19.9
	Operate TA-312 (CT09AB)	3	A: 56.9	CM7	58.9
				CM14	35.7
			B: 55.7	CM7	62.5
				CM14	34.8

Duty Area	HOPT Tasks	Steps	Mean	JKT Items	Mean
Communications (con't)					
	Check parts of TA-312 (CT10AB)	1	A: 95.5 B: 95.7	(none) (none)	—— ——

Duty Area	HOPT Tasks	Steps	Mean	JKT Items	Mean
First Aid	Administer mouth-to mouth resuscitation	7	A: 62.9 (FA01A)	FA6	46.7
				FA7	19.5
			B: no task	(none)	——
	CPR (FA02AB)	17	A: 40.5	FA1	23.9
				FA4	13.1
			B: 43.2	FA1	25.9
				FA4	15.3
	Treat for shock (FA03AB)	4	A: 48.8	FA2	29.9
				FA3	61.2
			B: 52.5	FA2	34.3
				FA3	63.7
	Fireman's carry	3	A: 62.4 (FA04A)	(none)	——
				(none)	——
			B: no tasks		
	Administer first aid for abdominal wound	8	A: 49.3 (FA05A)	FA5	21.9
			B: no task	(none)	——
	First aid for amputated limb	15	A: 49.2 (FA07A)	FA8	62.6
			B: no task	FA9	56.9
	Chest pressure-arm lift artificial respiration	7	A: no task	(none)	——
			B: 17.9 (FA08B)	FA13	52.0
				FA14	40.8
	Put on battle dressing	13	A: no task	(none)	——
			B: 61.4 (FA09B)	FA10	48.7
				FA11	43.0
	First aid sucking chest wound	7	A: no task	(none)	——
			B: 51.5 (FA10B)	FA12	30.6

	HOPT			JKT	
Duty Area	Tasks	Steps	Mean	Items	Mean
LAW (Light Antitank Weapon)					
	Restore expanded LAW (LA01AB)	4	A: 72.2	LAW4	60.8
				LAW9	52.4
			B: 72.0	LAW4	59.6
				LAW9	55.8
	Prepare to fire (LA2AB)	8	A: 56.9	LAW2	68.1
				LAW3	72.9
				LAW6	41.2
				LAW7	46.1
				LAW10	71.4
				LAW12	47.6
				LAW14	20.5
			B: 55.8	LAW2	69.8
				LAW3	72.7
				LAW6	36.5
				LAW7	47.1
				LAW10	72.8
				LAW12	50.1
				LAW14	18.3
	Take immediate action (LA03AB)	13	A: 41.4	LAW1	9.4
				LAW5	33.1
				LAW8	27.0
				LAW11	39.9
			B: 41.0	LAW1	7.7
				LAW5	33.3
				LAW8	25.0
				LAW11	36.3

Duty Area	HOPT Tasks	Steps	Mean	JKT Items	Mean
NBC	Give appropriate visual alarm	5	A: 90.5 (NB01A)	(none)	——
			B: no tasks	(note)	——
	Put on and wear protective clothing	30	A: 50.9 (NB02A)	NBC3A	32.5
				NBC7A	54.7
				NBC8A	48.8
				NBC9A	58.7
				NBC10A	81.7
			B: no	(none)	——
	Drink while masked	10	A: 71.8 (NB03A)	NBC5A	43.2
				NBC11A	54.5
			B: no tasks	(none)	——
	First aid for nerve gas casualty	13	A: 56.4 (NB04A)	NBC1A	20.1
				NBC2A	71.1
				NBC6A	60.6
			B: no tasks	(none)	——
	First aid for blistering agent	6	A: 51.7 (NB05A)	NBC4A	43.8
			B: no tasks	(none)	——
	Inspect/Maintain M17 mask	12	A: no tasks	(none)	——
			B: 66.5 (NB06B)	NBC1B	48.6
				NBC2B	43.2
				NBC3B	66.9
				NBC5B	44.1
	Identify NATO markers	7	A: no tasks	(none)	——
			B: 22.4 (NB07B)	NBC6B	57.4
				NBC7B	65.3
				NBC8B	31.0
	Remove mask	10	A: no tasks	(none)	——
			B: 41.5 (NB08B)	(none)	——
	React to aerial spray	29	A: no tasks	(none)	——
			B: 57.1 (NB09B)	NBC10B	55.1
	Treat choking agent casualty	9	A: no tasks	(none)	——
			B: 60.4 (NB10B)	NBC4B	41.1
				NBC9B	64.3
	Prepare NBC-1 report (NB13AB)	17	A: 60.1	(none)	——
			B: 60.1	(none)	——

Duty Area	HOPT Tasks	Steps	Mean	JKT Items	Mean
Security and Intelligence					
	Collect information	7	A: 51.8 (SI01A)	SI1	55.2
				SI2	54.2
				SI3	42.5
				SI4	41.1
				SI5	45.6
				SI6	47.2
				SI7	63.4
			B: no tasks	SI1	50.5
				SI2	55.9
				SI3	40.0
				SI4	38.8
				SI5	43.3
				SI6	40.2
				SI7	65.1
	Observe information	8	A: 59.5 (SI02A)	(none)	——
			B: no tasks	(none)	——
	Perform search and safeguard procedures (SI03AB)	4	A: 58.3	SI8	27.1
				SI9	42.2
				SI10	77.0
			B: 63.6	SI8	29.5
				SI9	38.0
				SI10	78.1
	Inspect and tag (SI04A)	10	A: 67.4	SI11	53.4
				SI12	19.6
	Inspect and tag (SI04B)	10	B: 67.6	SI11	51.7
				SI12	17.9
	Pass friendly personnel through lines	5	A: no task	(none)	——
			B: 49.4 (SI05B)	SI13	30.1
				SI14	45.9

Duty Area	HOPT Tasks	Steps	Mean	JKT Items	Mean
SAW					
	Inspect SAW: Visual inspection (SL01AB)	1	A: 60.2 B: 67.6	(none) (none)	—— ——
	Inspect SAW: Operations inspection (SL02AB)	17	A: 35.1 B: 41.7	SAW2 SAW3 SAW2 SAW3	35.9 52.7 35.9 52.5
	Maintain SAW: Fieldstrip (SL03AB)	37	A: 68.3 B: 71.3	SAW1 SAW1	71.5 70.7
	Maintain SAW: Assemble (SL04AB)	27	A: 74.1 B: 77.7	SAW4 SAW5 SAW6 SAW4 SAW5 SAW6	61.8 53.7 67.4 61.9 53.9 68.6

Duty Area	HOPT Tasks	Steps	Mean	JKT Items	Mean
Grenade Launcher					
	Operations inspection (GL01AB)	16	A: 62.5	GL8	69.4
			B: 62.7	GL8	69.4
	Prepare launcher for firing (GL02AB)	2	A: 98.3	GL1	51.5
				GL7	41.1
			B: 98.7	GL1	41.9
				GL7	41.5
	Confirm zero (GL04AB)	10	A: 44.0	GL4	60.7
				GL5	47.5
				GL9	31.4
				GL12	21.7
			B: 43.8	GL4	62.7
				GL5	47.2
				GL9	32.4
				GL12	19.7
	Maintain launcher (GL05AB)	13	A: 46.6	GL3	39.8
				GL6	54.0
				GL10	43.2
				GL11	78.8
			B: 48.1	GL3	36.3
				GL6	56.2
				GL10	43.0
				GL11	76.9
	Emplace stakes (GL06AB)	6	A: 19.2	GL2	26.7
			B: 18.5	GL2	27.8

	HOPT			JKT	
Duty Area	Tasks	Steps	Mean	Items	Mean
Hand Grenades	Engage targets with hand grenade (HG01AB)	6	A: 51.6	HG01	36.7
				HG02	61.1
			B: 49.9	HG01	38.3
				HG02	62.6

Duty Area	HOPT Tasks	Steps	Mean	JKT Items	Mean
Mines					
	Install Claymore mine	17	A: 45.9 (MI01A)	MI01A	39.8
				MI02A	55.0
				MI03A	23.1
				MI05A	54.2
			B: no tasks	(none)	——
	Recover Claymore mine	8	A: 58.8 (MI02A)	MI04A	59.2
			B: no tasks	(none)	——
	Install Claymore mine with tripwires (MI03B)	19	A: no tasks	(none)	——
			B: 21.1	MI01B	40.4
				MI02B	40.0
				MI04B	18.3
				MI06B	28.1
				MI07B	52.4
	Recover Claymore mine with tripwires (MI04B)	9	A: no tasks	(none)	——
			B: 30.5	MI03B	43.9
				MI05B	40.6

APPENDIX E:

RANKING OF ITEMS IN TERMS OF DISCREPANCIES--BOTH FORMS

APPENDIX E
RANKING OF ITEMS IN TERMS OF DISCREPANCIES—BOTH FORMS
Note: Negative values refer to JKT items easier than HOPT task

DISCREPANCY VALUE	ITEM LABEL	ITEM MEAN	TASK LABEL	HOPT MEAN	HOPT TASK NAME	TEST FORM	DUTY AREA
-50.2	CM0009	62.8	CR12AB	12.6	CONSTRUCT FIELD-EXPEDIENT ANT	A	C
-48.5	CM0009	60.8	CR12AB	12.3	CONSTRUCT FIELD-EXPEDIENT ANT	B	C
-42.9	NBC07B	65.3	NB07B	22.4	IDENTIFY NATO MARKERS	B	N
-39.2	LN0017	76.1	LN03AB	36.9	DET LOC'N BY MAP-TERR ASSOC	B	L
-35.1	CM0011	47.7	CR12AB	12.6	CONSTRUCT FIELD-EXPEDIENT ANT	A	C
-35	NBC06B	57.4	NB07B	22.4	IDENTIFY NATO MARKERS	B	N
-34.6	CM0011	46.9	CM12AB	12.3	CONSTRUCT FIELD-EXPEDIENT ANT	B	C
-34.1	FA0013	52	FA08B	17.9	CHEST PRESS-ARM LIFT ART RESP	B	F
-33.4	LN0017	75.3	LN03AB	41.9	DET. LOC'N BY MAP-TERR ASSOC	A	L
-32.2	GL0011	78.8	GL05AB	46.6	MAINTAIN LAUNCHER	A	G
-30.8	NBC010	81.7	NB02A	50.9	PUT ON & WEAR PROT CLOTHING	A	N
-29.4	LN0021	66.3	LN03AB	36.9	DET LOC'N BY MAP-TERR ASSOC	B	L
-29.2	LN0007	54.1	LN01A	24.9	SETTING AZIMUTH AT NIGHT	A	L
-28.8	GL0011	76.9	GL05AB	48.1	MAINTAIN LAUNCHER	B	G
-26.3	TM0010	79.6	TM01AB	53.3	ESTABLISH LANDING ZONE	A	T
-26.2	MI007B	47.3	MI03B	21.1	INSTALL CLAYMORE MINE W/TRIPW	B	M
-24.5	LN0007	51.9	LN01B	27.4	SETTING AZIMUTH AT NIGHT	B	L
-23.3	TM0010	76.8	TM01AB	53.5	ESTABLISH LANDING ZONE	B	T
-22.9	FA0014	40.8	FA08B	17.9	CHEST PRESS-ARM LIFT ART RESP	B	F
-22.8	LN0021	64.7	LN03AB	41.9	DET. LOC'N BY MAP-TERR ASSOC	A	L
-22.3	LN0018	59.2	LN03AB	36.9	DET LOC'N BY MAP-TERR ASSOC	B	L
-21.4	LF0003	71	LF01AB	49.6	BATTLESIGHT ZERO M16A2 RIFLE	B	R
-21.3	CM0012	33.6	CR12AB	12.3	CONSTRUCT FIELD-EXPEDIENT ANT	B	C
-21.1	LN0022	58	LN03AB	36.9	DET LOC'N BY MAP-TERR ASSOC	B	L

APPENDIX E
RANKING OF ITEMS IN TERMS OF DISCREPANCIES—BOTH FORMS
Note: Negative values refer to JKT items easier than HOPT task

DISCREPANCY VALUE	ITEM LABEL	ITEM MEAN	TASK LABEL	HOPT MEAN	HOPT TASK NAME	TEST FORM	DUTY AREA
-20.9	TM0028	59	TM14AB	38.1	CALL FOR & ADJUST INDIR FIRE	A	T
-20	CM0012	32.6	CR12AB	12.6	CONSTRUCT FIELD-EXPEDIENT ANT	A	C
-19.8	LN0013	65.5	LN06A	45.7	DETERMINE GRID COORDINATES	A	L
-19.4	TM0023	45.1	TM08AB	25.7	DIRECT HELICOPTER LANDING & TO	B	T
-19.3	TM0023	46.2	TM08AB	26.9	DIRECT HELICOPTER LANDING & TO	A	T
-19.3	MI001B	40.4	MI03B	21.1	INSTALL CLAYMORE MINE W/TRIPW	B	M
-18.9	GL0004	62.7	GL04AB	43.8	CONFIRM ZERO	B	G
-18.9	MI002B	40	MI03B	21.1	INSTALL CLAYMORE MINE W/TRIPW	B	M
-18.8	TM0028	58.4	TM14AB	39.6	CALL FOR & ADJUST INDIR FIRE	B	T
-18.7	SI0010	77	SI03AB	58.3	SEARCH AND SAFEGUARD PROCEDURE	A	I
-18.7	LF0003	68.8	LF01AB	50.1	BATTLESIGHT ZERO M16A2 RIFLE	A	R
-18.5	LN016A	64.2	LN06A	45.7	DETERMINE GRID COORDINATES	A	L
-17.6	SAW003	52.7	SL02AB	35.1	INSPECT SAW: OPERATIONS INSP	A	S
-17.3	LN0014	63	LN06A	45.7	DETERMINE GRID COORDINATES	A	L
-17	LAW010	72.8	LA02AB	55.8	PREPARE TO FIRE	B	A
-16.9	LAW003	72.7	LA02AB	55.8	PREPARE TO FIRE	B	A
-16.7	GL0004	60.7	GL04AB	44	CONFIRM ZERO	A	G
-16.1	CM0010	28.4	CR12AB	12.3	CONSTRUCT FIELD-EXPEDIENT ANT	B	C
-16	LAW003	72.9	LA02AB	56.9	PREPARE TO FIRE	A	A
-15.1	LN0013	69.5	LN06B	54.4	DETERMINE GRID COORDINATES	B	L
-14.7	NBC02A	71.1	NB04A	56.4	FA FOR NERVE GAS CASUALTY	A	N
-14.5	LAW010	71.4	LA02AB	56.9	PREPARE TO FIRE	A	A
-14.5	SI0010	78.1	SI03AB	63.6	SEARCH AND SAFEGUARD PROCEDURE	B	I
-14.2	TM0029	52.3	TM14AB	38.1	CALL FOR & ADJUST INDIR FIRE	A	T
-14	LAW002	69.8	LA02AB	55.8	PREPARE TO FIRE	B	A
-14	LN0015	59.7	LN06A	45.7	DETERMINE GRID COORDINATES	A	L
-14	LN001B	50.9	LN07B	36.9	DET. LOC'N BY RESECTION	B	L

APPENDIX E

RANKING OF ITEMS IN TERMS OF DISCREPANCIES—BOTH FORMS

Note: Negative values refer to JKT items easier than HOPT task

DISCREPANCY VALUE	ITEM LABEL	ITEM MEAN	TASK LABEL	HOPT MEAN	HOPT TASK NAME	TEST FORM	DUTY AREA
-13.5	CM0010	26.1	CR12AB	12.6	CONSTRUCT FIELD-EXPEDIENT ANT	A	C
-13.4	FA0008	62.6	FA07A	49.2	FA AMPUTATED LIMB	A	F
-13.4	MI003B	43.9	MI04B	30.5	RECOVER CLAYMORE MINE W/TRIPW	B	M
-13.1	LN0001	51.8	LN07A	38.7	DET. LOC'N BY RESECTION	A	L
-13	LN0016	67.4	LN06B	54.4	DETERMINE GRID COORDINATES	B	L
-12.7	HG0002	62.6	HG01AB	49.9	ENGAGE TARGETS WITH H GRENADE	B	H
-12.4	FA0003	61.2	FA03AB	48.8	TREAT FOR SHOCK	A	F
-11.9	LN0018	53.8	LN03AB	41.9	DET. LOC'N BY MAP-TERR ASSOC	A	L
-11.6	SI0007	63.4	SI01A	51.8	COLLECT INFORMATION	A	I
-11.2	FA0003	63.7	FA03AB	52.5	TREAT FPR SHOCK	B	F
-11.2	LAW002	68.1	LA02AB	56.9	PREPARE TO FIRE	A	A
-10.9	TM0029	50.5	TM14AB	39.6	CALL FOR & ADJUST INDIR FIRE	B	T
-10.8	LN0022	52.7	LN03AB	41.9	DET. LOC'N BY MAP-TERR ASSOC	A	L
-10.8	SAW003	52.5	SL02AB	41.7	INSPECT SAW: OPERATIONS INSP	B	S
-10.5	CM0004	73	CR01A	62.5	OPERATION INSPECTION AN/PRC-77	A	C
-10.3	LN0014	64.7	LN06B	54.4	DETERMINE GRID COORDINATES	B	L
-10.1	MI005B	40.6	MI04B	30.5	RECOVER CLAYMORE MINE W/TRIPW	B	M
-9.5	HG0002	61.1	HG01AB	51.6	ENGAGE TARGETS WITH H GRENADE	A	H
-9.3	GL0002	27.8	GL06AB	18.5	EMPLACE STAKES	B	G
-9.1	MI002A	55	MI02A	45.9	INSTALL CLAYMORE MINE	A	M
-8.6	NBC08B	31	NB07B	22.4	IDENTIFY NATO MARKERS	B	N
-8.6	CM0003	76	CR05B	67.4	TAKE IMMEDIATE ACTION	B	C
-8.5	TM0004	76.6	TL01AB	68.1	MOVE INDIVIDUALLY	B	T
-8.3	MI005A	54.2	MI05A	45.9	INSTALL CLAYMORE MINE	A	M
-8.1	GL0006	56.2	GL05AB	48.1	MAINTAIN LAUNCHER	B	G
-7.9	LN0008	32.8	LN01A	24.9	SETTING AZIMUTH AT NIGHT	A	L

APPENDIX E

RANKING OF ITEMS IN TERMS OF DISCREPANCIES—BOTH FORMS

Note: Negative values refer to JKT items easier than HOPT task

DISCREPANCY VALUE	ITEM LABEL	ITEM MEAN	TASK LABEL	HOPT MEAN	HOPT TASK NAME	TEST FORM	DUTY AREA
-7.8	NBC009	58.7	NB02A	50.9	PUT ON & WEAR PROT CLOTHING	A	N
-7.5	GL0002	26.7	GL06AB	19.2	EMPLACE STAKES	A	G
-7.4	GL0006	54	GL05AB	46.6	MAINTAIN LAUNCHER	A	G
-7.4	TM0025	45.5	TM14AB	38.1	CALL FOR & ADJUST INDIR FIRE	A	T
-7	LN0008	34.4	LN01B	27.4	SETTING AZIMUTH AT NIGHT	B	L
-7	MI006B	28.1	MI03B	21.1	INSTALL CLAYMORE MINE W/TRIPW	B	M
-6.9	GL0008	69.4	GL01AB	62.5	OPERATIONS INSPECTION	A	G
-6.8	CM0007	62.5	CT09AB	55.7	OPERATE TA-312	B	C
-6.7	GL0008	69.4	GL01AB	62.7	OPERATIONS INSPECTION	B	G
-6.2	LN0015	60.6	LN06B	54.4	DETERMINE GRID COORDINATES	B	L
-4.8	TM0025	44.4	TM14AB	39.6	CALL FOR & ADJUST INDIR FIRE	B	T
-4.2	NBC06A	60.6	NB04A	56.4	FA FOR NERVE GAS CASUALTY	A	N
-3.9	NBC09B	64.3	NB10B	60.4	TREAT CHOKING AGENT CASUALTY	B	N
-3.8	NBC07A	54.7	NB02A	50.9	PUT ON & WEAR PROT CLOTHING	A	N
-3.6	TM0004	73.7	TL01AB	70.1	MOVE INDIVIDUALLY	A	T
-3.5	GL0005	47.5	GL04AB	44	CONFIRM ZERO	A	G
-3.4	SI0001	55.2	SI01A	51.8	COLLECT INFORMATION	A	I
-3.4	GL0005	47.2	GL04AB	43.8	CONFIRM ZERO	B	G
-3.2	SAW001	71.5	SL03AB	68.3	MAINTAIN SAW: FIELDSTRIP	A	S
-3.1	LN0023	40	LN03AB	36.9	DET LOC'N BY MAP-TERR ASSOC	B	L
-2.7	CM0013	21.1	CT08AB	18.4	REPAIR WIRE OF TA-312	A	C
-2.4	SI0002	54.2	SI01A	51.8	COLLECT INFORMATION	A	I
-2	CM0007	58.9	CT09AB	56.9	OPERATE TA-312	A	C
-1.2	LN0020	38.1	LN03AB	36.9	DET LOC'N BY MAP-TERR ASSOC	B	L
-1.1	TM0027	39.2	TM14AB	38.1	CALL FOR & ADJUST INDIR FIRE	A	T

APPENDIX E
RANKING OF ITEMS IN TERMS OF DISCREPANCIES—BOTH FORMS
Note: Negative values refer to JKT items easier than HOPT task

DISCREPANCY VALUE	ITEM LABEL	ITEM MEAN	TASK LABEL	HOPT MEAN	HOPT TASK NAME	TEST FORM	DUTY AREA
-1	CM0013	19.9	CT08AB	18.9	REPAIR WIRE OF TA-312	B	C
-0.8	SAW002	35.9	SL02AB	35.1	INSPECT SAW: OPERATIONS INSP	A	S
-0.7	TM0031	38.8	TM14AB	38.1	CALL FOR & ADJUST INDIR FIRE	A	T
-0.4	MI004A	59.2	MI04A	58.8	RECOVER CLAYMORE MINE	A	M
-0.4	NBC03B	66.9	NB06B	66.5	INSPECT/MAINTAIN M17 MASK	B	N
0.4	LN0005	38.3	LN07A	38.7	DET. LOC'N BY RESECTION	A	L
0.6	SAW001	70.7	SL03AB	71.3	MAINTAIN SAW: FIELDSTRIP	B	S
1	LN0009	58.7	LN02AB	59.7	PACE DISTANCE	A	L
1.5	LAW011	39.9	LA03AB	41.4	TAKE IMMEDIATE ACTION	A	A
1.7	TM0031	37.9	TM14AB	39.6	CALL FOR & ADJUST INDIR FIRE	B	T
1.9	LN005B	35	LN07B	36.9	DETERMINE LOC'N BY RESECTION	B	L
2	NBC10B	55.1	NB09B	57.1	REACT TO AERIAL SPRAY	B	N
2.1	NBC08A	48.8	NB02A	50.9	PUT ON & WEAR PROT CLOTHING	A	N
2.8	CM0001	59.7	CR01A	62.5	OPERATION INSPECTION AN/PRC-77	A	C
2.8	MI004B	18.3	MI03B	21.1	INSTALL CLAYMORE MINE W/TRIPW	B	M
2.9	GL0001	51.5	GL02AB	54.4	PREPARE LAUNCHER FOR FIRING	A	G
3.1	LN0009	55.9	LN02AB	59	SETTING AZIMUTH AT NIGHT	B	L
3.3	LN006B	24.1	LN01B	27.4	SETTING AZIMUTH AT NIGHT	B	L
3.4	GL0010	43.2	GL05AB	46.6	MAINTAIN LAUNCHER	A	G
3.5	SI0014	45.9	SI05B	49.4	PASS FRIENDLY PERSONNEL THRU	B	I
4.4	LN0023	37.5	LN03AB	41.9	DET. LOC'N BY MAP-TERR ASSOC	A	L
4.5	LN0019	32.4	LN03AB	36.9	DET LOC'N BY MAP-TERR ASSOC	B	L
4.6	LN0006	20.3	LN01A	24.9	SETTING AZIMUTH AT NIGHT	A	L
4.6	SI0006	47.2	SI01A	51.8	COLLECT INFORMATION	A	I
4.7	LN0003	51.5	LN08A	56.2	DET. LOC'N BY INTERSECTION	A	L
4.7	LAW011	36.3	LA03AB	41	TAKE IMMEDIATE ACTION	B	A

APPENDIX E
RANKING OF ITEMS IN TERMS OF DISCREPANCIES—BOTH FORMS
Note: Negative values refer to JKT items easier than HOPT task

DISCREPANCY VALUE	ITEM LABEL	ITEM MEAN	TASK LABEL	HOPT MEAN	HOPT TASK NAME	TEST FORM	DUTY AREA
5.1	GL0010	43	GL05AB	48.1	MAINTAIN LAUNCHER	B	G
5.5	TM0027	34.1	TM14AB	39.6	CALL FOR & ADJUST INDIR FIRE	B	T
5.7	LAW012	50.1	LA02AB	55.8	PREPARE TO FIRE	B	A
5.8	SAW002	35.9	SL02AB	41.7	INSPECT SAW: OPERATIONS INSP	B	S
5.9	CM0002	77	CR04AB	82.9	ASSEMBLE RADIO AN/PRC-77	A	C
6.1	MI001A	39.8	MI01A	45.9	INSTALL CLAYMORE MINE	A	M
6.2	SI0005	45.6	SI01A	51.8	COLLECT INFORMATION	A	I
6.7	SAW006	67.4	SL04AB	74.1	MAINTAIN SAW: ASSEMBLE	A	S
6.8	LN0019	35.1	LN03AB	41.9	DET. LOC'N BY MAP-TERR ASSOC	A	L
6.8	GL0003	39.8	GL05AB	46.6	MAINTAIN LAUNCHER	A	G
7	LN0020	34.9	LN03AB	41.9	DET. LOC'N BY MAP-TERR ASSOC	A	L
7.6	TM0008	60.5	TL01AB	68.1	MOVE INDIVIDUALLY	B	T
7.7	LAW005	33.3	LA03AB	41	TAKE IMMEDIATE ACTION	B	A
7.9	TM0007	60.2	TL01AB	68.1	MOVE INDIVIDUALLY	B	T
7.9	NBC04A	43.8	NB05A	51.7	FA FOR BLISTERING AGENT	A	N
8.3	LAW005	33.1	LA03AB	41.4	TAKE IMMEDIATE ACTION	A	A
8.5	TM0015	68.3	TM09AB	76.8	CONTROL UNIT MOVMNT WHEN NOT	A	T
8.7	LAW007	47.1	LA02AB	55.8	PREPARE TO FIRE	B	A
9.1	SAW006	68.6	SL04A	77.7	MAINTAIN SAW: ASSEMBLE	B	S
9.3	LAW012	47.6	LA02AB	56.9	PREPARE TO FIRE	A	A
9.3	SI0003	42.5	SI01A	51.8	COLLECT INFORMATION	A	I
9.5	LN0004	46.7	LN08A	56.2	DET. LOC'N BY INTERSECTION	A	L
9.6	TM0009	58.5	TL01AB	68.1	MOVE INDIVIDUALLY	B	T
10.2	TM0008	59.9	TL01AB	70.1	MOVE INDIVIDUALLY	A	T
10.5	CM0002	75.4	CR04AB	85.9	ASSEMBLE RADIO AN/PRC-77	B	C

APPENDIX E
RANKING OF ITEMS IN TERMS OF DISCREPANCIES—BOTH FORMS
Note: Negative values refer to JKT items easier than HOPT task

DISCREPANCY VALUE	ITEM LABEL	ITEM MEAN	TASK LABEL	HOPT MEAN	HOPT TASK NAME	TEST FORM	DUTY AREA
10.6	TM0015	66.1	TM09AB	76.7	CONTROL UNIT MOVMNT WHEN NOT	B	T
10.7	SI0004	41.1	SI01A	51.8	COLLECT INFORMATION	A	I
10.7	CM0001	57.5	CR01B	68.2	OPERATION INSPEC OF AN/PRC-77	B	C
10.8	LAW007	46.1	LA02AB	56.9	PREPARE TO FIRE	A	A
11	TM0033	28.6	TM14AB	39.6	CALL FOR & ADJUST INDIR FIRE	B	T
11.3	CM0015	34.6	CT07AB	45.9	INSTALL TELEPHONE SET TA-312	B	C
11.3	TM0033	26.8	TM14AB	38.1	CALL FOR & ADJUST INDIR FIRE	A	T
11.4	GL0009	32.4	GL04AB	43.8	CONFIRM ZERO	B	G
11.4	LAW004	60.8	LA01AB	72.2	RESTORE EXPANDED LAW	A	A
11.6	CM0015	34.3	CT07AB	45.9	INSTALL TELEPHONE TA-312	A	C
11.6	HG0001	38.3	HG01AB	49.9	ENGAGE TARGETS WITH H GRENADE	B	H
11.8	GL0003	36.3	GL05AB	48.1	MAINTAIN LAUNCHER	B	G
12.2	TM0007	57.9	TL01AB	70.1	MOVE INDIVIDUALLY	A	T
12.3	SAW004	61.8	SL04AB	74.1	MAINTAIN SAW: ASSEMBLE	A	S
12.4	TM0009	57.7	TL01AB	70.1	MOVE INDIVIDUALLY	A	T
12.4	LAW004	59.6	LA01AB	72	RESTORE EXPANDED LAW	B	A
12.6	GL0009	31.4	GL04AB	44	CONFIRM ZERO	A	G
12.7	FA0010	48.7	FA09B	61.4	PUT ON BATTLE DRESSING	B	F
12.9	NV002	52.4	NV02AB	65.3	OPERATIONS INSPECTION	A	V
13.2	NV004A	38.7	NV03A	51.9	CLEAN COMPONENTS	A	V
13.5	TM0035	40	TM01AB	53.5	ESTABLISH LANDING ZONE	B	T
14	SI0011	53.4	SI04A	67.4	INSPECT & TAG	A	I
14.2	NV003A	37.7	NV03A	51.9	CLEAN COMPONENTS	A	V
14.4	LAW008	27	LA03AB	41.4	TAKE IMMEDIATE ACTION	A	A
14.6	TM0003	59.8	TL05AB	74.4	CAMOUFLAGE SELF AND EQUIP	A	T
14.9	HG0001	36.7	HG01AB	51.6	ENGAGE TARGETS WITH H GRENADE	A	H

APPENDIX E
RANKING OF ITEMS IN TERMS OF DISCREPANCIES—BOTH FORMS
Note: Negative values refer to JKT items easier than HOPT task

DISCREPANCY VALUE	ITEM LABEL	ITEM MEAN	TASK LABEL	HOPT MEAN	HOPT TASK NAME	TEST FORM	DUTY AREA
15.1	TM0001	60.9	TL05AB	76	CAMOUFLAGE SELF & EQUIPMENT	B	T
15.5	TM0032	24.1	TM14AB	39.6	CALL FOR & ADJUST INDIR FIRE	B	T
15.7	LAW006	41.2	LA02AB	56.9	PREPARE TO FIRE	A	A
15.8	SAW004	61.9	SL04AB	77.7	MAINTAIN SAW: ASSEMBLE	B	S
15.9	SI0011	51.7	SI04B	67.6	INSPECT AND TAG	B	I
16	LAW008	25	LA03AB	41	TAKE IMMEDIATE ACTION	B	A
16.1	SI0009	42.2	SI03AB	58.3	SEARCH AND SAFEGUARD PROCEDURE	A	I
16.2	FA0006	46.7	FA01A	62.9	MOUTH-TO-MOUTH RESUSCITATION	A	F
16.2	LAW009	55.8	LA01AB	72	RESTORE EXPANDED LAW	B	A
16.5	CM0005	40.7	CR02A	57.2	VISUAL INSPECTION OF AN/PRC-77	A	C
16.5	TM0003	59.5	TL05AB	76	CAMOUFLAGE SELF & EQUIPMENT	B	T
16.6	FA0001	23.9	FA02AB	40.5	CPR	A	F
16.9	TM0030	59.8	TM09AB	76.7	CONTROL UNIT MOVMNT WHEN NOT	B	T
17	TM0001	57.4	TL05AB	74.4	CAMOUFLAGE SELF AND EQUIP	A	T
17.1	TM0005	57.3	TL05AB	74.4	CAMOUFLAGE SELF AND EQUIP	A	T
17.3	FA0001	25.9	FA02AB	43.2	CPR	B	F
17.3	NBC011	54.5	NB03A	71.8	DRINK WHILE MASKED	A	N
17.4	TM0035	35.9	TM01AB	53.3	ESTABLISH LANDING ZONE	A	T
17.8	CM0008	28.1	CT07AB	45.9	INSTALL TELEPHONE SET TA-312	B	C
17.9	NBC01B	48.6	NB06B	66.5	INSPECT/MAINTAIN M17 MASK	B	N
18	TM0032	20.1	TM14AB	38.1	CALL FOR & ADJUST INDIR FIRE	A	T
18.2	FA0002	34.3	FA03AB	52.5	TREAT FOR SHOCK	B	F
18.3	NV0002	47.3	NV02AB	65.6	OPERATIONS INSPECTION	B	V
18.4	NBC03A	32.5	NB02A	50.9	PUT ON & WEAR PROT CLOTHING	A	N
18.4	FA0011	43	FA09B	61.4	PUT ON BATTLE DRESSING	B	F
18.9	FA0002	29.9	FA03AB	48.8	TREAT FOR SHOCK	A	F

APPENDIX E
RANKING OF ITEMS IN TERMS OF DISCREPANCIES—BOTH FORMS
Note: Negative values refer to JKT items easier than HOPT task

DISCREPANCY VALUE	ITEM LABEL	ITEM MEAN	TASK LABEL	HOPT MEAN	HOPT TASK NAME	TEST FORM	DUTY AREA
19.3	LAW006	36.5	LA02AB	55.8	PREPARE TO FIRE	B	A
19.3	NBC04B	41.1	NB10B	60.4	TREAT CHOKING AGENT CASUALTY	B	N
19.3	CM0008	26.6	CT07AB	45.9	INSTALL TELEPHONE TA-312	A	C
19.3	SI0013	30.1	SI05B	49.4	PASS FRIENDLY PERSONNEL THRU	B	I
19.8	LAW009	52.4	LA01AB	72.2	RESTORE EXPANDED LAW	A	A
19.8	TM0030	57	TM09AB	76.8	CONTROL UNIT MOVMNT WHEN NOT	A	T
20.4	SAW005	53.7	SL04AB	74.1	MAINTAIN SAW: ASSEMBLE	A	S
20.9	FA0012	30.6	FA10B	51.5	FIRST AID SUCKING CHEST WND	B	F
20.9	CM0014	34.8	CT09AB	55.7	OPERATE TA-312	B	C
21.2	CM0014	35.7	CT09AB	56.9	OPERATE TA-312	A	C
21.8	TM0019	55	TM09AB	76.8	CONTROL UNIT MOVMNT WHEN NOT	A	T
22	LN0011	34.3	LN11AB	56.3	MEASURE DISTANCE ON A MAP	B	L
22.1	TM0005	53.9	TL05AB	76	CAMOUFLAGE SELF & EQUIPMENT	B	T
22.3	GL0012	21.7	GL04AB	44	CONFIRM ZERO	A	G
22.4	NBC05B	44.1	NB06B	66.5	INSPECT/MAINTAIN M17 MASK	B	N
22.8	MI003A	23.1	MI03A	45.9	INSTALL CLAYMORE MINE	A	M
23.3	NBC02B	43.2	NB06B	66.5	INSPECT/MAINTAIN M17 MASK	B	N
23.8	SAW005	53.9	SL04AB	77.7	MAINTAIN SAW: ASSEMBLE	B	S
24.1	GL0012	19.7	GL04AB	43.8	CONFIRM ZERO	B	G
25.2	TM0014	28.3	TM01AB	53.5	ESTABLISH LANDING ZONE	B	T
25.6	SI0009	38	SI03AB	63.6	SEARCH AND SAFEGUARD PROCEDURE	B	I
26.2	TM0014	27.1	TM01AB	53.3	ESTABLISH LANDING ZONE	A	T
27	LN0011	31.2	LN11AB	58.2	MEASURE DISTANCE ON A MAP	A	L
27.2	TM0012	26.3	TM01AB	53.5	ESTABLISH LANDING ZONE	B	T
27.3	TM0019	49.4	TM09AB	76.7	CONTROL UNIT MOVMNT WHEN NOT	B	T
27.4	FA0004	13.1	FA02AB	40.5	CPR	A	F
27.4	FA0005	21.9	FA05AB	49.3	FA ABDOMINAL WOUND	A	F

APPENDIX E
RANKING OF ITEMS IN TERMS OF DISCREPANCIES—BOTH FORMS
Note: Negative values refer to JKT items easier than HOPT task

DISCREPANCY VALUE	ITEM LABEL	ITEM MEAN	TASK LABEL	HOPT MEAN	HOPT TASK NAME	TEST FORM	DUTY AREA
27.5	TM0013	26	TM01AB	53.5	ESTABLISH LANDING ZONE	B	T
27.8	TM0012	25.5	TM01AB	53.3	ESTABLISH LANDING ZONE	A	T
27.9	FA0004	15.3	FA02AB	43.2	CPR	B	F
28.6	NBC05A	43.2	NB03A	71.8	DRINK WHILE MASKED	A	N
28.9	TM0013	24.4	TM01AB	53.3	ESTABLISH LANDING ZONE	A	T
30.3	LN0012	26	LN11AB	56.3	MEASURE DISTANCE ON A MAP	B	L
31.2	SI0008	27.1	SI03AB	58.3	SEARCH AND SAFEGUARD PROCEDURE	A	I
31.6	LN0012	26.6	LN11AB	58.2	MEASURE DISTANCE ON A MAP	A	L
32	TM0011	21.3	TM01AB	53.3	ESTABLISH LANDING ZONE	A	T
32	LAW001	9.4	LA03AB	41.4	TAKE IMMEDIATE ACTION	A	A
32.3	LN0002	23.9	LN08A	56.2	DET. LOC'N BY INTERSECTION	A	L
33.3	LAW001	7.7	LA03AB	41	TAKE IMMEDIATE ACTION	B	A
33.9	TM0011	19.6	TM01AB	53.5	ESTABLISH LANDING ZONE	B	T
34.1	SI0008	29.5	SI03AB	63.6	SEARCH AND SAFEGUARD PROCEDURE	B	I
35.3	LN0010	21	LN11AB	56.3	MEASURE DISTANCE ON A MAP	B	L
35.5	LN0024	19.8	LN04AB	55.3	DET. AZIMUTH FROM 1 PT TO ANO	A	L
36.2	TM0002	38.2	TL05AB	74.4	CAMOUFLAGE SELF AND EQUIP	A	T
36.3	NBC01A	20.1	NB04A	56.4	FA FOR NERVE GAS CASUALTY	A	N
36.4	LAW014	20.5	LA02AB	56.9	PREPARE TO FIRE	A	A
37.5	LAW014	18.3	LA02AB	55.8	PREPARE TO FIRE	B	A
37.7	LN0010	20.5	LN11AB	58.2	MEASURE DISTANCE ON A MAP	A	L
38.1	TM0002	37.9	TL05AB	76	CAMOUFLAGE SELF & EQUIPMENT	B	T
39.8	LN0024	15.7	LN04AB	55.5	DET. AZIMUTH FROM 1 PT TO ANO	B	L
40.9	TM0006	33.5	TL05AB	74.4	CAMOUFLAGE SELF AND EQUIP	A	T

APPENDIX E

RANKING OF ITEMS IN TERMS OF DISCREPANCIES—BOTH FORMS

Note: Negative values refer to JKT items easier than HOPT task

DISCREPANCY VALUE	ITEM LABEL	ITEM MEAN	TASK LABEL	HOPT MEAN	HOPT TASK NAME	TEST FORM	DUTY AREA
41.3	NV001A	26.6	NV01A	67.9	VISUAL INSPECTION	A	V
42.3	TM0006	33.7	TL05AB	76	CAMOUFLAGE SELF & EQUIPMENT	B	T
43.4	FA0007	19.5	FA01A	62.9	MOUTH-TO-MOUTH RESUSCITATION	A	F
46.1	TM0016	30.7	TM09AB	76.8	CONTROL UNIT MOVMNT WHEN NOT	A	T
46.1	TM0020	30.6	TM09AB	76.7	CONTROL UNIT MOVMNT WHEN NOT	B	T
47	TM0020	29.8	TM09AB	76.8	CONTROL UNIT MOVMNT WHEN NOT	A	T
47.5	TM0018	29.2	TM09AB	76.7	CONTROL UNIT MOVMNT WHEN NOT	B	T
47.8	SI0012	19.6	SI04A	67.4	INSPECT & TAG	A	I
49.2	TM0018	27.6	TM09AB	76.8	CONTROL UNIT MOVMNT WHEN NOT	A	T
49.3	TM0022	27.5	TM09AB	76.8	CONTROL UNIT MOVMNT WHEN NOT	A	T
49.3	TM0022	27.4	TM09AB	76.7	CONTROL UNIT MOVMNT WHEN NOT	B	T
49.7	SI0012	17.9	SI04B	67.6	INSPECT AND TAG	B	I
50.5	TM0016	26.2	TM09AB	76.7	CONTROL UNIT MOVMNT WHEN NOT	B	T
51.6	TM0021	25.2	TM09AB	76.8	CONTROL UNIT MOVMNT WHEN NOT	A	T
52.5	TM0021	24.2	TM09AB	76.7	CONTROL UNIT MOVMNT WHEN NOT	B	T
52.7	TM0034	24.1	TM09AB	76.8	CONTROL UNIT MOVMNT WHEN NOT	A	T
53.9	TM0034	22.8	TM09AB	76.7	CONTROL UNIT MOVMNT WHEN NOT	B	T
56.8	GL0001	41.9	GL02AB	98.7	PREPARE LAUNCHER FOR FIRING	B	G
57.2	GL0007	41.1	GL02AB	98.3	PREPARE LAUNCHER FOR FIRING	A	G
57.2	GL0007	41.5	GL02AB	98.7	PREPARE LANCHER FOR FIRING	B	G
63.2	TM0017	13.5	TM09AB	76.7	CONTROL UNIT MOVMNT WHEN NOT	B	T
63.8	TM0017	13	TM09AB	76.8	CONTROL UNIT MOVMNT WHEN NOT	A	T

APPENDIX F:

LIST OF ITEMS WITH BELOW-CHANCE LEVELS OF CORRECT RESPONSES

APPENDIX F

LIST OF ITEMS WITH BELOW-CHANCE LEVELS OF CORRECT RESPONSES

Item name	Form A		Form B		
	Item mean	HOPT mean	Item mean	HOPT mean	Correct response
Setting Azimuth at Night(LN06)	20.3	24.9	24.1	27.4	C
Determine Azimuth One Pt to Another (LN24)	19.8	55.3	15.7	55.5	D
Determine Location by Intersection (LN02)	23.9	56.2	27.1	****	A
Measure Distance on a Map (LN10)	20.5	58.2	21.0	56.3	A
Establish Landing Zone (TM11)	21.3	53.3	19.6	53.5	C
Establish Landing Zone (TM13)	24.4	53.3	26.0	53.5	A
Control Unit Movement When Not in Contact (TM17)	13.0	76.8	13.5	76.7	D
Control Unit Movement When Not in Contact (TM21)	25.2	76.8	24.2	76.7	D
Control Unit Movement When Not in Contact (TM34)	24.1	76.8	22.8	76.7	D
Call for and Adjust Indirect Fire(TM32)	20.1	38.1	24.1	39.6	C
Repair Wire of TA-312 (CM13)	21.1	18.4	19.9	18.9	A
Administer Mouth-to-Mouth Resuscitation(FA7)	19.5	62.9	****	****	D
CPR (FA1)	23.9	40.5	25.9	43.2	C
CPR (FA4)	13.1	40.5	15.3	43.2	D

APPENDIX F: (Continued)

Item name	Form A		Form B		
	Item mean	HOPT mean	Item mean	HOPT mean	Correct response
Administer First Aid for Abdominal Wound(FA5)	21.9	49.3	****	****	A
Prepare to Fire (LAW14)	20.5	56.9	18.3	55.8	A
Take Immediate Action (LAW1)	9.4	41.4	7.7	41.0	D
Drink While Masked (NBC1A)	20.1	56.4	****	****	C
Inspect and Tag (SI12)	19.6	67.4	17.9	67.6	D
Confirm Zero(GL12)	21.7	44.0	19.7	43.8	D
Install Claymore Mine with Trip-wires(MI04B)	****	****	18.3	21.1	C

NOTE: Asterisks indicate that the task or item was not part of the test for that form.

APPENDIX G:

SCORESHEET FOR COMMUNICATIONS TASK 12, "CONSTRUCTING A FIELD-EXPEDIENT ANTENNA"

APPENDIX G

SCORESHEET FOR COMMUNICATIONS TASK 12, "CONSTRUCTING A FIELD-EXPEDIENT ANTENNA"

Say: This test covers your ability to construct a field-expedient antenna. You have before you (indicate ground cloth with equipment) all the equipment you should need. Here is your assigned frequency (3 x 5 card with 46.90 frequency). You are to construct a 1/2 wave omni-directional VHF antenna. Do you have any questions about these instructions? Begin.

NOTE TO SCORER: Check Marine's answer to step 1. If answer is wrong, say: Use 5 feet.

PERFORMANCE STEPS

	GO	NO-GO
1. Used frequency reference chart and formula to determine correct length of wire.	_____	_____
2. Cut antenna wire to 5 feet.	_____	_____
3. Stripped approximately 3/4 inch of wire.	_____	_____
4. Twisted the field wire antenna.	_____	_____
5. Attached the bare ends of the antenna to the antenna connection of the radio by screwing the antenna base over the leads into the antenna mounting hole.	_____	_____
6. Selected appropriate insulator (non-conductive).	_____	_____
7. Tied antenna wire to one end of the insulator.	_____	_____
8. Tied rope to other end of insulator.	_____	_____
9. Threw rope over tree limb and raised antenna until it was vertical.	_____	_____
10. Stripped about 2 inches off of each end of the ground wire.	_____	_____
11. Cut 2-3 feet of wire for the ground.	_____	_____
12. Stripped about 2 inches off of each end of the ground wire.		
13. Attached one end of the ground to any metal part of the radio.	_____	_____
14. Drove metal stake into ground near radio.	_____	_____
15. Attached other end of ground to metal stake.	_____	_____

APPENDIX H:

HOPT SCORESHEET FOR CONTROL UNIT MOVEMENT WHEN NOT IN CONTACT

APPENDIX H

HOPT SCORESHEET FOR CONTROL UNIT MOVEMENT WHEN NOT IN CONTACT

Say: This test covers your ability to control unit movement when not in contact. You are a squad leader located at grid point 277 529 (point). You must move to grid point 213 548 (point). The enemy is located in the high ground northwest of the river (point).

Enemy contact is *not likely* from your present location (point) to within small arms range of the town (point). When you are within small arms range enemy contact changes to *possible*. When you come within small arms range of the treeline of Hill 437 (point) enemy contact changes to *contact expected*. Do you have any questions?

NOTE TO SCORER: Repeat elements of scenario if asked.

PERFORMANCE STEPS	GO	NO-GO
Say: Indicate your general direction of movement with this pointer.		
1. Marine indicated a northwesterly route.	_____	_____

NOTE TO SCORER: Now lay the three white chips on the predesignated spots.

Say: Your squad, represented by this chip, is at this location (point to grid 229 530). What formation should it be in?

2. Indicated a good formation.	_____	_____

Good Formation	Poor Formation
Tactical column	Echelon (R or L)
Wedge	Line
	Vee

Say: Your squad is now here (point to grid 224 539). What formation should it be in?

3. Indicated a good formation. _____ _____

Good Formation	Poor Formation
Echelon right	Tactical column
Wedge	Line
	Vee
	Echelon left

Say: Your squad is now here (point to grid 217 544). What formation should it be in?

4. Indicated a good formation. _____ _____

Good Formation	Poor Formation
Line	Column
Vee	Echelon (R or L)
	Wedge

NOTE TO SCORER: Remove the three white chips from the TACWAR board and hand to Marine.

Say: Each white chip represents a fireteam of your squad. Assume the enemy is in this direction (point). Show me a *wedge* formation, using the chips on the board, and demonstrate the hand and arm signal.

5. Indicated a *wedge* formation. _____ _____

6. Gave the proper hand and arm signal for a wedge.

Say: Show me a column formation and demonstrate the hand and arm signal.

7. Indicated a *column* formation. _____ _____

8. Gave the proper hand and arm signal for a column. _____ _____

Say: An echelon left.

9. Indicated an *echelon left* formation. _____ _____

10.	Gave the proper hand and arm signal for an echelon left.	_____	_____
	Say: A vee formation.		
11.	Indicated a *vee* formation.	_____	_____
12.	Gave the proper hand and arm signal for a vee.	_____	_____
	Say: A line formation.		
13.	Indicated a *line* formation.	_____	_____
14.	Gave the proper hand and arm signal for a line.	_____	_____

APPENDIX I:

ITEM QUALITY CONTROL FOR JOB-KNOWLEDGE TESTS

APPENDIX I

ITEM QUALITY CONTROL FOR JOB-KNOWLEDGE TESTS

PART I--Determining the Appropriate Test Mode

	Yes	Maybe	No
Is the task appropriately measured by means of a multiple-choice item? (i.e., is the task knowledge-dependent, reading-dependent, and time-independent?)	___	___	___
Is the task more appropriately measured by means of a fill-in, alternate choice, multiple true-false, essay, or testlet format?	___	___	___
Does the task require physical coordination, reflex responses, complex construction, or complex reasoning?	___	___	
Is the item appropriate for the skill level of the Marines to be tested?	___	___	___
Does the item require a skill other than what it is intended to measure? (e.g., reading skills, understanding of notation, memory for facts that can be easily looked up in a manual)?	___	___	___
Does the item require less than what it is intended to measure? (e.g., does it unnecessarily simplify a complex task into a choice between alternatives?)	___	___	___

PART II--Item technical review (modified from Hambleton, 1980)

Test Item Characteristics (Mark "/" for Yes, "X" for No and "?" for unsure)	Test Item Numbers		
	1	2	3
Is the item stem clearly written for the intended Marines?			
Is the stem free of irrelevant material?			
Is a single problem clearly defined in the item stem?			
Are the answer choices clearly written for the intended group of Marines?			
Are the answer choices free of irrelevant material?			
Is there a correct answer or a *clearly* best answer?			
Have words like "always", "none", or "all" been removed?			
Are likely student mistakes used to prepare incorrect choices?			
Is "all of the above" avoided as an answer choice?			
Are the answer choices arranged in a logical sequence (if one exists)?			
Was the correct answer randomly positioned among the available choices?			
Are all repetitious words or expressions removed from the answer choices and included in the item stem?			
Are all the answer choices of the same length?			
Do the item stem and answer choices follow standard rules of punctuation and grammar?			
Are all negatives underlined?			
Are grammatical cues between the item stem and the answer choices, which might give the correct answer away, removed?			

	Test Item Numbers		
	1	2	3
Are letters used in front of the possible answers to identify them?			
Have expressions like "which of the following is *not*" been avoided?			
Disregarding any technical flaws that might exist in the test item, how well do you think the content of the test item matches with the duty area of the content defined by the domain specification? (1=poor, 2=fair, 3=good, 4=very good, 5=excellent)			